Available Online through
www.ijpbs.com (or) www.ijpbsonline.com

IJPBS |Volume 3| Issue 3 |JUL-SEP|2013|72-87

Research Article
Biological Sciences

MOLECULAR MODELING AND COMPARITIVE STUDY OF DIAMINOPALMITIC ACID OF MICROBIAL TUBERCULOSIS IN H37RV

A Sushma*[1], M.G Shambhu[1], Kusum paul[1]

1Department of Biotechnology, The Oxford college of Engineering, Bommanahalli, Hosur road, Bangalore – 560 068, Karnataka state, India.

***Corresponding Author Email:** laksh.nov6@gmail.com

ABSTRACT

The increasing emergence of multiple drug resistant TB (MDR-TB) and extensively drug-resistant TB poses a serious threat to the control of tuberculosis disease. Furthermore, it is reported that 79% of MDR- TB cases are 'super strains'. Here is the Aspartyl beta-semi aldehyde dehydrogenase (ASADH) is an important enzyme, occupying the first branch position of the biosynthetic pathway of the aspartate family of amino acids in bacteria, fungi and higher plants. It catalyses reversible dephosphorylation of L: -beta-aspartyl phosphate (betaAP) to L: -aspartate-beta-semialdehyde (ASA), a key intermediate in the biosynthesis of diaminopimelic acid (DAP)-an essential component of cross linkages in bacterial cell walls.

. It has been found that mtASADH exhibits structural features common to bacterial ASADH, while other structural motifs are not present. Structural analysis of various domains in mtASADH reveals structural conservation among all bacterial ASADH proteins. The results suggest that the probable mechanism of action of the mtASADH enzyme might be same as that of other bacterial ASADH. Analysis of the structure of mtASADH will shed light on its mechanism of action and may help in designing suitable antagonists against this enzyme that could control the growth of Mycobacterium tuberculosis.

KEY WORDS

Tuberculosis, ASADH, H37Rv, Mycobacterium tuberculosis bacteria, Molecular Docking

INTRODUCTION

Tuberculosis, MTB, or TB is a bacterial infection that can spread through the lymph nodes and bloodstream to any organ in the body (short for tubercle bacillus) is a common, and in many cases lethal, infectious disease caused by various strains of mycobacterium, usually Mycobacterium tuberculosis.

Tuberculosis typically attacks the lungs, but can also affect other parts of the body. Most people who are exposed to TB never develop symptoms, because the bacteria can live in an inactive form in the body. But if the immune system weakens, such as in people with HIV or adults, TB bacteria can become active. In their active state, TB bacteria cause death of tissue in the organs they infect. Active TB disease can be fatal if left untreated**[1].**

One third of the world's population is thought to have been infected with M. tuberculosis, with new infections occurring at a rate of about one per second. In 2007, there were an estimated 13.7 million chronic active cases globally, while in 2010, there were an estimated 8.8 million new cases and 1.5 million associated deaths, mostly occurring in developing countries.

About 80% of the population in many Asian and African countries test positive in tuberculin tests, while only 5–10% of the United States population tests positive. More people in the developing world contract tuberculosis because of compromised

immunity, largely due to high rates of HIV infection and the corresponding development of AIDS **[2].**

MATERIALS AND METHODS

- **UNIPROTKB (**URL:http://www.uniprot.org/**)**

The UniProt Knowledgebase(UniProtKB)is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation**[3].**

- **PROTEIN DATA BANK(PDB):**URL: http://www.rcsb.org/pdb/

The Protein Data Bank (PDB) is a repository for the 3-D structural data of large biological molecules, such as proteins and nucleic acids**[4].**

- **PUBCHEM:**URL: http://www.ncbi.nlm.nih.gov/pccompound/

PubChem is a database of chemical molecules and their activities against biological assays. The system is maintained by the National Center for Biotechnology Information (NCBI), a component of the National Library of Medicine**[5].**

- **CHEMSPIDER:**URL: http://cssp.chemspider.com/

ChemSpider is a free chemical database. With over 28 million unique chemicals on the database linked out to over 400 data sources the platform provides access to experimental and predicted data (properties, spectra etc.), links to publications, patents and a myriad of other resources**[6].**

- **BLAST: (**URL: http://blast.ncbi.nlm.nih.gov/Blast.cgi)

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences **[7].**

- **LALIGN:**

Lalign is a program designed by EBI tools .It compares two protein /DNA sequences for local similarity and shows the local sequence. LALIGN and PALIGN compare two sequences to identify local sequence similarities.

- **PROTPARAM:**URL: http://web.expasy.org/protparam/

Protparam (References / Documentation) is a tool which allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered sequence **[8].**

- **CLUSTALW:**URL: http://www.genome.jp/tools/clustalw/

ClustalW2 is a general purpose multiple sequence alignment program for DNA or proteins. It attempts to calculate the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen **[9].**

- **SOPMA:**URL: http://npsa-pbil.ibcp.fr/cgi-bin/NPSA/npsa_sopma.html

A new method called the self-optimized prediction method (SOPM) has been described for the prediction of the secondary structure of proteins **[10].**

- **SWISS MODEL:**URL: http://swissmodel.expasy.org/

SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server or from the program Deep View (Swiss Pdb-Viewer) **[11].**

- **SAVS:**URL: http://nihserver.mbi.ucla.edu/SAVES/

Structural Analysis and Verification ServerChecks the stereo chemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry.

- **QSITE FINDER:**URL: http://www.modelling.leeds.ac.uk/qsitefinder/

Q-Site Finder is a new method of ligand binding site prediction. It works by binding hydrophobic (CH3) probes to the protein, and finding clusters of probes with the most favourable binding energy [**12].**

- **SPDBV**

Swiss-PdbViewer (SPdbV) is an easy-to-use and powerful molecular modeling program. In addition to its many built in features, it is tightly linked to Swiss-Model (http://www.expasy.ch/swissmod/SWISS-MODEL.html), an automated homology modeling server run by the Geneva Biomedical Research Center.

- **CHEMSKETCH:**

ChemSketch is an all-purpose chemical drawing and graphics software. Use templates or free-hand

- **HYPERCHEM:**

HyperChem is a sophisticated molecular modeling environment that is known for its quality, flexibility, and ease of use.

❖ **AUTO DOCK:**

Auto Dock is a suite of automated docking tools. It is designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure

❖ **MOLEGRO:**

Molegro Virtual Docker is an integrated platform for predicting protein - ligand interactions.

❖ **PYMOL:**

Pymol is a user-sponsored molecular visualization system on an open-source foundation.

SEQUENCE ANALYSIS

In bioinformatics, the term **sequence analysis** refers to the process of subjecting a DNA, RNA or peptide sequence to any of a wide range of analytical methods to understand its features, function, structure, or evolution.

METHODOLOGIES USED IN SEQUENCE ANALYSIS

a. Dynamic programming, b. Artificial Neural Network, Hidden Markov Model,d. Support Vector Machine,e. Clustering,f. Bayesian Network,g. Regression Analysis.

Here initially sequence analysis of ASADH in mycobacterium tuberculosis and ASADH in h37rv has been performed. It has been performed by local alignment. Local alignment has been performed to deduce the local similarity between two sequences I.e. between ASADH in mycobacterium tuberculosis and the same in H37RV using Lalign program.

ALIGNMENT

In bioinformatics, a sequence alignment is a way of arranging the sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences **[13].**

REPRESENTATION

Alignments are commonly represented both graphically and in text format. In almost all sequence alignment representations, sequences are written in rows arranged so that aligned residues appear in successive columns. Systematic representation of sequence alignment is shown in the **Figure 1**[14].

PAIRWISE ALIGNMENT

Pair wise Sequence Alignment is used to identify regions of similarity that may indicate functional, structural and/or evolutionary relationships between two biological sequences**[15](**protein or nucleic acid). Pair wise sequence alignment methods are used to find the best-matching piecewise (local) or global alignments of two query sequences **[16].**

When ASADH OF Mycobacterium tuberculosis and the same in h37rv subjected for local alignment, Function Lalign finds 33.8% identity and 66.2% similarity in 68aa overlaps i.e. from 43-105 residues and 249-312 residues and 41.2% identity (70.6 similar) in 17aa overlap from 62-78:312-328)and also **36.7**% identity(60.0 similar) in 30aa overlap(225-249:298-327).Thus Lalign checks for local similarity between two sequences. This is accomplished to know the similarities between ASADH in mycobacterium tuberculosis and ASADH in H37Rv

SEQUENCE SIMILARITY SEARCH

To accomplish sequence similarity the most popular blast has been used. The search provides list of sequences similar to query sequence

BLAST (BASIC LOCAL ALIGNMENT SEARCH TOOL)

When the query sequence i.e. ASADH (target protein) of mycobacterium tuberculosis is subjected for BLAST it results in over 100 similar sequences.

PROTEIN STRUCTURE PREDICTION

PRIMARY STRUCTURE PREDICTION

❖ **PROTPARAM**

Protparam computes various physico-chemical properties that can be deduced from a protein sequence. The parameters computed by ProtParam include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).

SECONDARY STRUCTURE PREDICTION

Secondary structure prediction is a set of techniques in bioinformatics that aim to predict the secondary structures of proteins and nucleic acid sequences based only on knowledge of their primary structure

This is an important step in drug discovery wherein here the target protein ASADH is subjected to SOPMA tool for secondary structure prediction. The SOPMA results showed that there are alpha helices in 37.10%, extended coil in 21.74%, beta turn in 8.99% and random coil in 32.17%.

HOMOLOGY MODELING

Homology modeling, also known as comparative modeling of protein, refers to constructing an atomic-resolution model of the "target" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "template").

STEPS IN HOMOLOGY MODELING

1. Target identification,2. Template identification,3. Alignment, 4. Build loop (Back bone chain),5. Scan Loop (side chain),6. Energy Minimization,7. Validation

COMPOUNDS SELECTION

Tetrahydroquinolines moiety is an important structural feature of various natural products and pharmaceutical agents that have exhibiting a broad spectrum of biological activities. Several of these compounds are naturally occurring.

Substituted tetrahydroquinolines are the core structures in many important pharmacological agents and drug molecules such as antiulcer, anti rhythmic and cardiovascular agents, anticancer drugs, immunosuppressant, and as high affinity ligands at the glycine site of the NMDA receptors.

Chemists have made substantial contributions in the design and development of nucleic acid cleavage agents for use as structural probes and therapeutic agents. In this photodynamic therapy (PDT) is an emerging method of non-invasive treatment of cancer in which drugs shows localized toxicity on photo activation at the tumor cells leaving the healthy cells un affected. Importantly, the type and the efficiency of the photo cleavage reaction will depend on the binding site that the photo nuclease occupies.

Further to here carried out of three component condensation between Alkynes, aniline & benzaldehyde, which resulted in a simple preparation of some substituted tetrahydroquinolines**.**

CHEMICAL STRUCTURE DESIGN OF COMPOUNDS

Chemical structure design of compounds is performed to optimize the physicochemical properties of their compounds and to explore property-based structure optimization.

ENERGY MINIMIZATION OF COMPOUNDS:

This is accomplished by making use of ChemSketch wherein description about it is explained briefly above in materials and methods.

COMPUTATION OF QSAR PROPERTIES

Quantitative structure-activity relationship models are regression models used in the classification models used in the chemical and biological sciences and engineering.

DOCKING STUDIES

In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex.

RESULTS AND DISCUSSION

SEQUENCE SIMILARITY BETWEEN ASADH PROTEIN OF MYCOBACTERIUM TUBERCULOSIS AND ASADH OF H37RV

Here the Aspartyl beta-semialdehyde dehydrogenase (ASADH) is an important enzyme, occupying the first branch position of the biosynthetic pathway of the aspartate family of amino acids in bacteria, fungi and higher plants.. Since the aspartate pathway is unique to plants and bacteria, and ASADH is the key enzyme in this pathway, it becomes an attractive target for antimicrobial agent development.

Here initially sequence analysis of ASADH in mycobacterium tuberculosis and ASADH in h37rv has been performed. It has been performed by local alignment. Local alignment has been performed to deduce the local similarity between two sequences I.e. between ASADH in mycobacterium tuberculosis and the same in H37RV using Lalign program. Protein sequence of ASADH in mycobacterium tuberculosis retrieved from UNIPROTKB and downloaded in FASTA format with the ID.The query sequence of ASADH protein of mycobacterium tuberculosis was consisting of 345 residues is shown in the **Figure 2**.

The longest sequence of ASADH with Accession No: was selected for 3D model development that contains 345 amino acid residues with molecular weight.

SEQUENCE SIMILARITY SEARCH

Psi-blast is performed to find the templates for the target protein ASADH and the result of it is shown in the **Figure 3.**

PRIMARY STRUCTURE PREDICTION

The Primary structure of the query protein ASADH is predicted using program called ExPASy Protparam Tool. For the predicted aminoacid composition refer

Table 1. For the atomic composition of target protein refer **Table 2**.Predicted parameters and the corresponding values are shown in the following listing

Formula: $C_{1595}H_{2579}N_{453}O_{494}S_7$

Total number of atoms: 5128

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Total number of negatively charged residues (Asp + Glu): 43

Total number of positively charged residues (Arg + Lys): 32

Extinction coefficients: Ext. coefficient 10095Abs 0.1% (=1 g/l) 0.279, assuming all pairs of Cysteine residues form Cysteine

Ext. coefficient: 9970Abs 0.1% (=1 g/l) 0.275, assuming all Cysteine residues are reduced

Estimated half-life: The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).

>20 hours (yeast, in vivo).

>10 hours (Escherichia coli, in vivo).

Instability index: The instability index (II) is computed to be 29.02

This classifies the protein as stable.

Aliphatic index: 99.54

SECONDARY STRUCTURE PREDICTION

Secondary structure of the ASADH is found using SOPMA to determine the composition of secondary structures and its result is shown in the **Figure 4 And Figure 5**.

TEMPLATE IDENTIFICATION

For the query sequence ASADH protein of Mycobacterium tuberculosis PSI-BLAST was performed to get the similar sequences wherein 500 Hits were found. since it is required to find the better templates having PDB structure Multiple sequence alignment was performed and the results is shown in the **Figure 6** .The structural similarity is more important than sequence similarity and here the templates 3TZ6_A, 3VOS_A Tb strains were exhibiting more structural similarities with the target protein ASADH and hence they are retrieved from NCBI database. **Table 3** illustrates their definition, no of residues, geneId and accession no respectively.

HOMOLOGY MODELING

Homology Model of ASADH () was constructed by using Swiss Model program. After aligning query with templates **3TZ6_A, 3VOS_A,** that alignment file is given as input to Swiss model program. PDB Structure of Target [ASADH] protein taken from SPDBV IS shown in the **Figure 7**.And the alignment of target and template is shown in the **Figure 8**.

This alignment is saved as a file named aligned. Sequence and the structure is saved as protein alignment.pdb and this file is given as input to Program Swiss model to get the 3D structure.For the 3D-Model of aligned protein refer **Figure 9**.

And the graph is obtained which shows the estimated absolute quality of the model and it is shown in the **Figure 10**. The QMEAN4 score of the whole model reflects the predicted model reliability ranging from 0 to 1. Here since the z-score QMEAN is 0.61 the model is said to have a good quality [19].

HOMOLOGY MODELING OF ASPARTYL BETA-SEMIALDEHYDE DEHYDROGENASE PROTEIN

The result of alignment was employed to build new homology model. Reliability of new homology model for ASADH was identified by Ramachandran Plot. After the optimization and energy minimization process, the best model was selected from 3D models generated for ASADH protein on the basis of Swiss Model. Foe the Ramachandran plot of the modeled ASADH protein refer **Figure 11**.

The first Ramachandran plot for the model shows the amino acids outside the cavities and they are ARG147, ARG 39, ALA 330.This happens when the protein structure gets damaged or if the internal energy is high due to the presence of Proline or glycine near those amino acids. The second Ramachandran plot shows the modeled plot where in the amino acids outside the cavities are sent to inside the cavities by building loop, scanning loop and energy minimization. Energy minimization of three dimensional structures is vital for providing the maximum stability to the protein and energy minimization of the predicted model is -7162. Ramachandran Plot drawn through Swiss Pdb Viewer program.

ACTIVE SITEPREDICTION

Once we find the target structure then it is required to find active sites in that.The Ligand binding site of ASADH protein is predicted.The result of Q-site finder shows that predicted binding site cavity volume modeled of Asadh protein is 160 cubic angstroms and the coordinates of the binding box around predicted site has minimum coordinates (247, 31, 68) and has maximum coordinates are (261, 48, 83) it as shown in the **Figure 12**.For the active site amino acids refer **Table 4**.

CHEMICALSTRUCTURE DESIGN OF COMPOUNDS

Compounds are designed using chemsketch and for the designed compounds refer **Figure 13**.

ENERGY MINIMIZATION OF INDIVIDUAL COMPOUNDSCOMPUTED FROM HYPERCHEM

Energy minimization or optimization is important in molecular docking calculations. It routinely optimizes geometries within the binding site. Complex optimization allows the ligand to obtain minimum energy pose within the active site cavity of the protein. It also allows the relaxation of protein to certain extent which can account for the conformational changes that happen in the protein structure on binding of the ligands.

In addition the calculated energies can also be used to estimate the binding energy which help in quantifying the binding process and have better understanding of the molecular recognition. The free energy of binding is the change in free energy that occurs on binding, ΔG binding= G complex - G separated where G complex and G separated are the free energies of the Complexed and no interacting protein and ligand respectively. In order to represent the salvation, the solvent molecules have been replaced with dielectric medium. Parameterized molecular mechanics force field (MMFF) has been used for the optimization.The energy minimization of sample compounds and the values are included in the **Table 5.**

COMPUTATION OF QSAR PROPERTIES

Quantitative structure–activity relationship models (QSAR models) are regression or classification models used in the chemical and biological sciences and engineering.For the QSAR properties computed from HyperChem refer **Table 6.**

RESULTS OF DOCKING STUDY

Docking is the simulation of a candidate ligand binding to a receptor .It depends on the search algorithm and scoring function. Here by using Auto dock.Docking poses the best conformation of ten sample substituted compounds named as S1,S2,S3,S4,S5,S6,S7,S8,S9 and S10 in the binding sites of modeled ASADH protein wherein the conformations are shown and the Conformation for compound I-3 is shown in the **Figure 14**.

Here the third compound named I-3 poses which is subjected for docking posses four hydrogen bond interactions and from that it is known that it can be serve as one of the good drug candidate.The Conformation for compound I-9 is shown in the **Figure 15**.Here the ninth compound named I-9 which is subjected for docking posses 4 hydrogen bond interactions and thus from this can be known that it can serve as one of the good drug candidate.

The Conformation for compound I-10 is shown in the **Figure 16**

Here the tenth compound named I-10 which is subjected for docking posses 6 hydrogen bond interactions and thus from this can be known that it can serve as one of the good drug candidate.

ANALYSIS OF DOCKING STUDIES

The **Table 7** consists of values of docking results from which analysis can be done. It says that,

The third sample compound named as I-3 is involved in four hydrogen bond interactions with the receptor molecule ASADH and the bonds connected to GLY12(glycine) and SER37(serine) is having low total energies of -9.37729 and -11.5533 respectively, and from this it can be concluded that this compound can serve as an affective ligand.

The ninth sample compound named as I-9 is involved in four hydrogen bond interactions with the receptor molecule ASADH and the bonds connected to THR281(Threonine) and VAL111(valine) is having low total energies of -13.6434 and -9.54715 respectively, and from this it can be concluded that this compound can serve as an affective ligand.

Also The tenth sample compound named as I-10 is involved in six hydrogen bond interactions with the receptor molecule ASADH and the bond connected to

SER71(serine) is having low total energies of -9.07825, and from this it can be concluded that this compound also can serve as an affective ligand.

Further submitting these compounds to chemical formulation can come with the effective.

CONCLUSION

The comparative Molecular modeling of aspartyl beta-semialdehyde dehydrogenase of Mycobacterium tuberculosis and H37Rv reveals that they exhibit common structures. Thus from the analysis the ligands exhibiting higher bonding interactions can be considered as compounds that effectively bind to the active site of the target.

These hydrogen bonds are connected to the amino acids having lower energies. Out of ten compounds taken for docking the compounds named as I-3,I-9 and I-10 and can serve as effective ligands for drug discovery process. Also The tenth sample compound named as I-10 is involved in six hydrogen bond interactions with the receptor molecule ASADH and the bond connected to SER71(serine) is having low total energies of -9.07825, and from this it can be concluded that this compound also can serve as the best ligand.

REFERENCES

[1]Niemann S et al. (2002). "Mycobacterium africanum Subtype II Is Associated with Two Distinct Genotypes and Is a Major Cause of Human Tuberculosis in Kampala, Uganda". Journal of Clinical Microbiology40 (9): 3398–405.

[2]Niobe-Eyangoh SN et al. (2003). "Genetic Biodiversity of Mycobacterium tuberculosis Complex Strains from Patients with Pulmonary Tuberculosis in Cameroon". Journal of Clinical Microbiology41 (6): 2547–53. doi:10.1128/JCM.41.6.2547-2553.2003

[3]Magrane M. and the UniProt consortium UniProt Knowledgebase: a hub of integrated protein data Database, 2011: bar009 (2011).

[4]Berman, H. M. (January 2008). "The Protein Data Bank: a historical perspective". *Acta Crystallographica Section A: Foundations of Crystallography* A64 (1): 88–95. doi:10.1107/S0108767307035623. PMID 18156675.

[5]"PubChem and the American Chemical Society". *Reshaping Scholarly Communication*. USA: University of California.

[6]Antony John Williams (Jan-Feb 2008). "ChemSpider and Its Expanding Web: Building a Structure-Centric Community for Chemists". *Chemistry International* 30 (1).

[7]Altschul, Stephen; Gish, Warren; Miller, Webb; Myers, Eugene; Lipman, David (1990). "Basic local alignment search tool". *Journal of Molecular Biology* 215 (3): 403–410. doi:10.1016/S0022-2836(05)80360-2. PMID 2231712.

[8]. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D.,BairochA.;(In) John M. Walker (ed): The Proteomics Protocols Handbook, HumanaPress(2005).pp. 571-607

[9]Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003). "Multiple sequence alignment with the Clustal series of programs". *Nucleic Acids Res* 31 (13): 3497–3500. doi:10.1093/nar/gkg500. PMC 168907. PMID 12824352

[10]Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 1995 Dec;11(6):681-684

[11]Arnold K., Bordoli L., Kopp J., and Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 22,195-201.

[12] Laurie AT, Jackson RM (2005). Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics,* 21: 1908-1916 Pubmed

[13]Mount DM. (2004). *Bioinformatics: Sequence and Genome Analysis* (2nd ed.). Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. ISBN 0-87969-608-7.

[14].ClustalW2FAQs http://www.ebi.ac.uk/Tools/clustalw2/help.html#color

[15]Mount DM. (2004). *Bioinformatics: Sequence and Genome Analysis* (2nd ed.). Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. ISBN 0-87969-608-7

[16]Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, Batzoglou S (2003). "Glocal alignment: finding rearrangements during alignment". *Bioinformatics*. 19. Suppl 1 (90001): i54–62. doi:10.1093/bioinformatics/btg1005. PMID 12855437.

```
AAB24882        TYHMCQFHCRYVNNHSGEKLYECNERSKAFSCPSHLQCHKRRQIGEKTHEHNQCGKAFPT 60
AAB24881        --------------------YECNQCGKAFAQHSSLKCHYRTHIGEKPYECNQCGKAFSK 40
                                     ****: .***:  * *:** * :****.:* *******..

AAB24882        PSHLQYHERTHTGEKPYECHQCGQAFKKCSLLQRHKRTHTGEKPYE-CNQCGKAFAQ- 116
AAB24881        HSHLQCHKRTHTGEKPYECNQCGKAFSQHGLLQRHKRTHTGEKPYMNVINMVKPLHNS 98
                 **** *:***********:***:**.: .***************    :  *.: :
```

Figure 1

P0A542 (DHAS_MYCTU) Reviewed, UniProtKB/Swiss-Prot
>sp|P0A542|DHAS_MYCTUAspartate-semialdehydedehydrogenase OS=Mycobacterium tuberculosis GN=asd PE=1 SV=1
MGLSIGIVGATGQVGQVMRTLLDERDFPASAVRFFASARSQGRKLAFRGQEIEVEDAETA
DPSGLDIALFSAGSAMSKVQAPRFAAAGVTVIDNSSAWRKDPDVPLVVSEVNFERDAHRR
PKGIIANPNCTTMAAMPVLKVLHDEARLVRLVVSSYQAVSGSGLAGVAELAEQARAVIGG
AEQLVYDGGALEFPPPNTYVAPIAFNVVPLAGSLVDDGSGETDEDQKLRFESRKILGIPD
LLVSGTCVRVPVFTGHSLSINAEFAQPLSPERARELLDGATGVQLVDVPTPLAAAGVDES
LVGRIRRDPGVPDGRGLALFVSGDNLRKGAALNTIQIAELLTADL

Figure 2

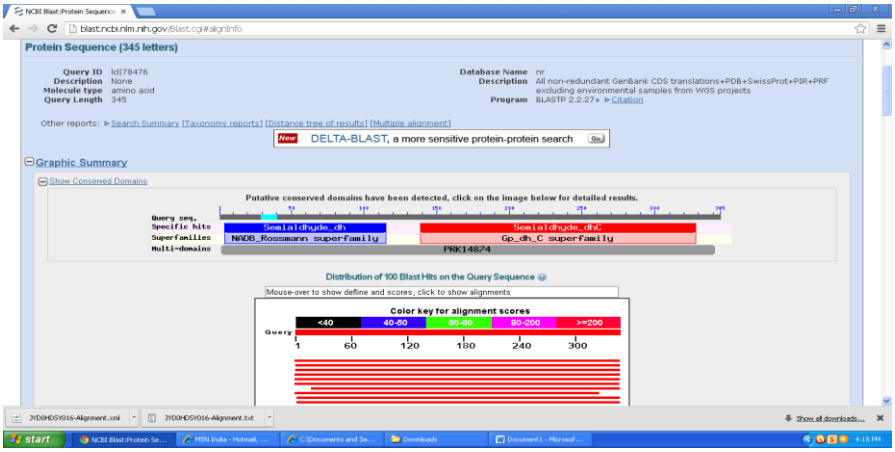


Figure 3

```
        10        20        30        40        50        60        70
        |         |         |         |         |         |         |
MGLSIGIVGATGQVGQVMRTLLDERDFPASAVRFFASARSQGRKLAFRGQEIEVEDAETADPSGLDIALF
tteeeeeeecthhhhhhhhhhhttccccchheeeehccccttceeeettcceeehhcchhhhhhceeeee
SAGSAMSKVQAPRFAAAGVTVIDNSSAWRKDPDVPLVVSEVNFERDAHRRPKGIIANPNCTTMAAMPVLK
ecccchhhhhhhhhhhttceeeeecccccccttccceeeecccchhhhhhcceeeecttcchhhhhhhhh
VLHDEARLVRLVVSSYQAVSGSGLAGVAELAEQARAVIGGAEQLVYDGGALEFPPPNTYVAPIAFNVVPL
hhhhhhhhhheeehhhhhcccccchhhhhhhhhhhhhhhtchheeeetcccccccccccccccceeeecc
AGSLVDDGSGETDEDQKLRFESRKILGIPDLLVSGTCVRVPVFTGHSLSINAEFAQPLSPERARELLDGA
ccccccttcccchhhhhhhhhhhhhhccccceeecceeeeeeecccchheeehhccccchhhhhhhhhtt
TGVQLVDVPTPLAAAGVDESLVGRIRRDPGVPDGRGLALFVSGDNLRKGAALNTIQIAELLTADL
ttceeeeccccccttcchheeeehccccccccttteeeeeecchhctthhhhhhhhhhhhhhhh

Sequence length :   345

SOPMA :
   Alpha helix     (Hh) :   128 is  37.10%
   3(10) helix     (Gg) :     0 is   0.00%
   Pi helix        (Ii) :     0 is   0.00%
   Beta bridge     (Bb) :     0 is   0.00%
   Extended strand (Ee) :    75 is  21.74%
   Beta turn       (Tt) :    31 is   8.99%
   Bend region     (Ss) :     0 is   0.00%
   Random coil     (Cc) :   111 is  32.17%
   Ambigous states (?)  :     0 is   0.00%
   Other states         :     0 is   0.00%
```

Figure 4

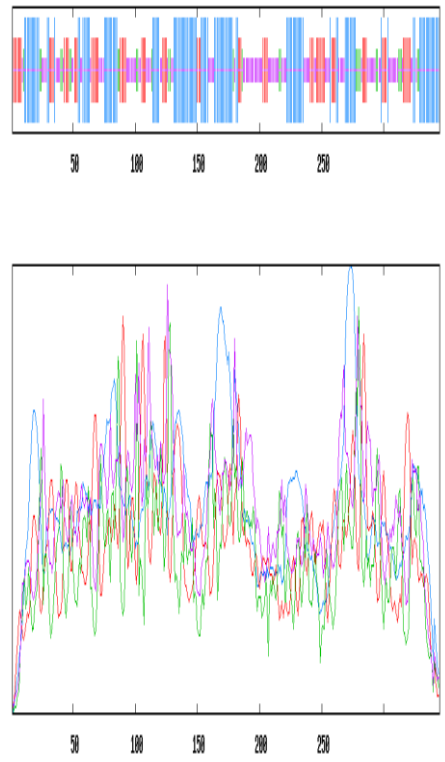


Figure 5

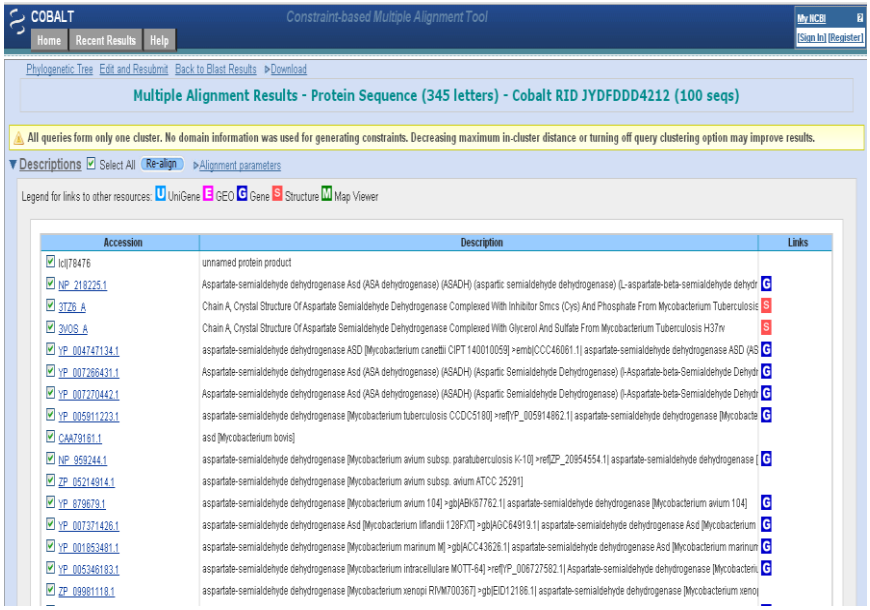


Figure 6

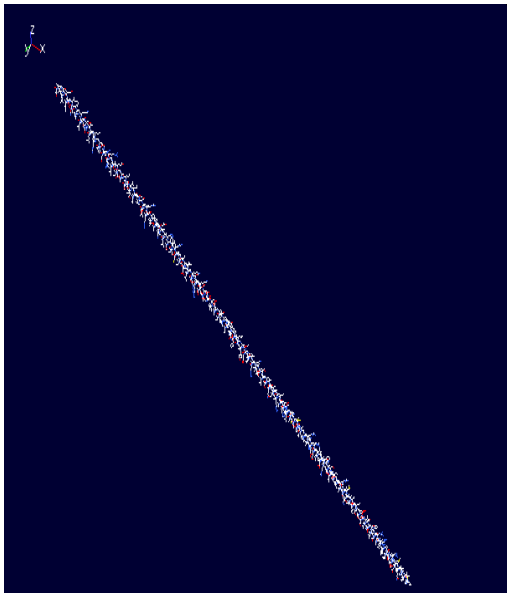


Figure 7

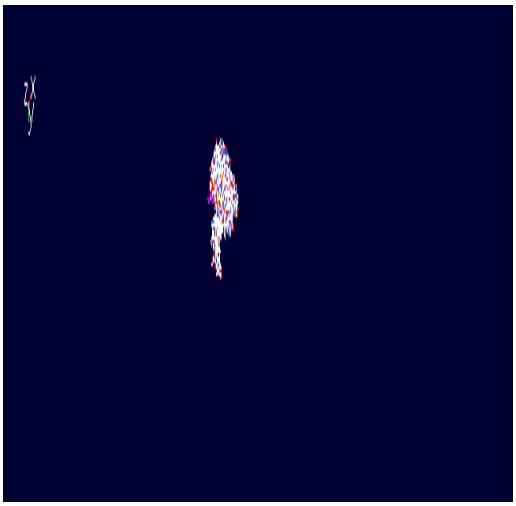


Figure 8

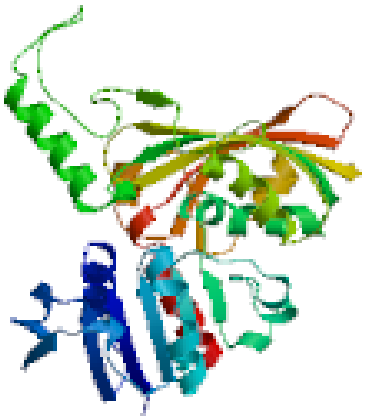


Figure 9

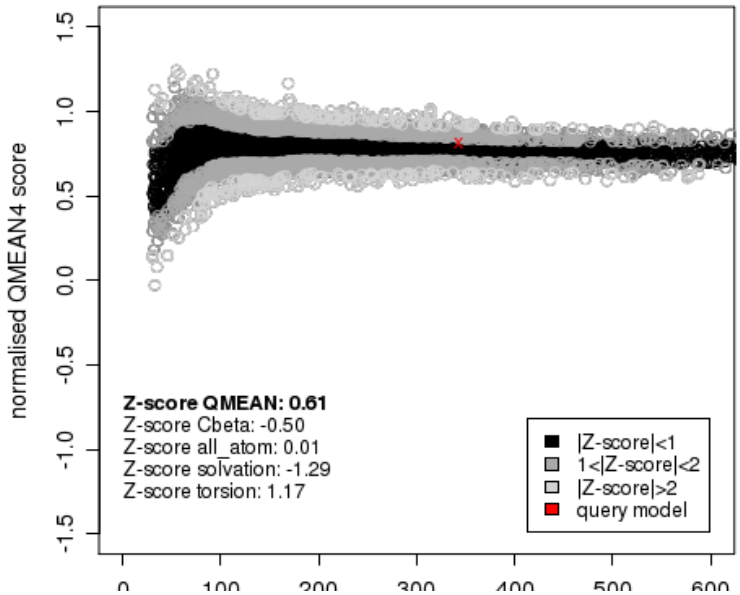
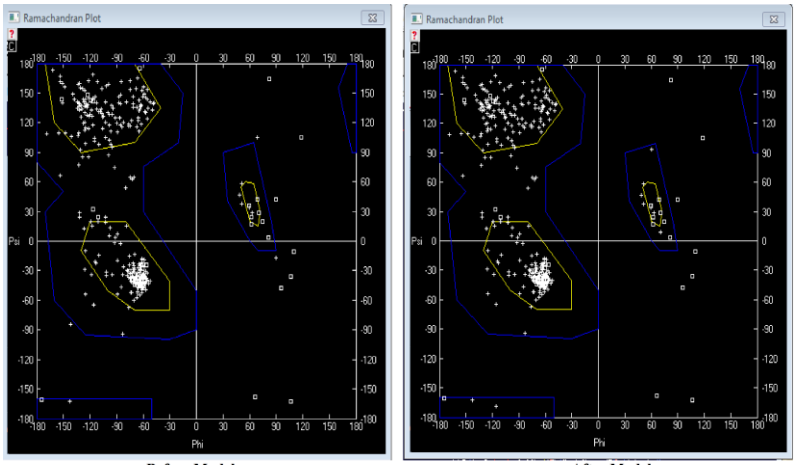


Figure 10



Before Model After Model

Figure 11

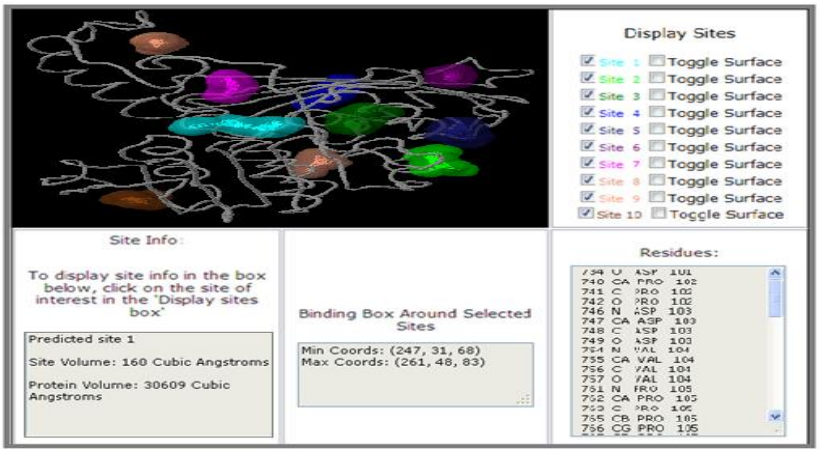


Figure 12

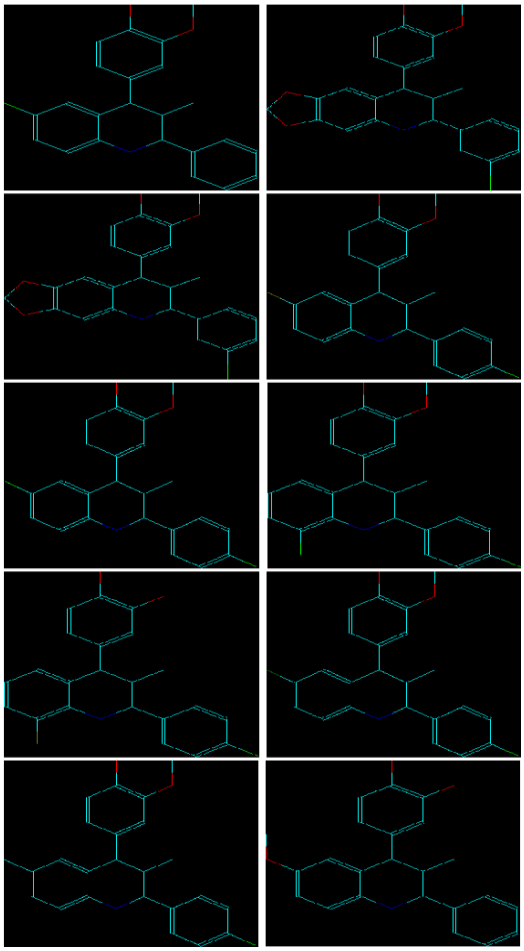


Figure 13

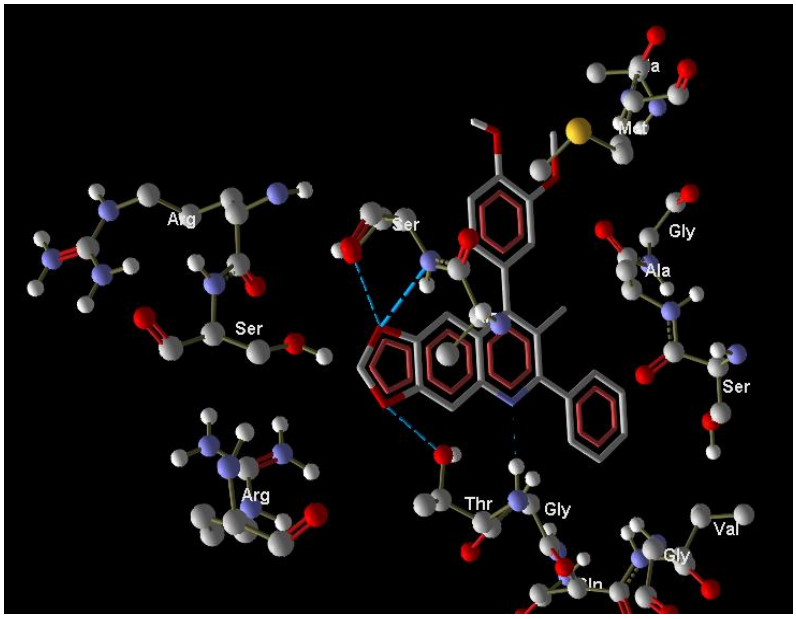


Figure 14

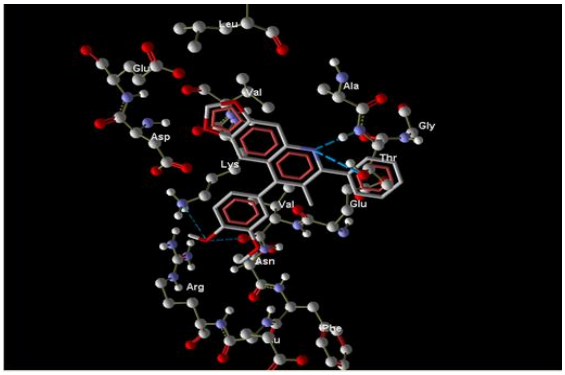


Figure 15

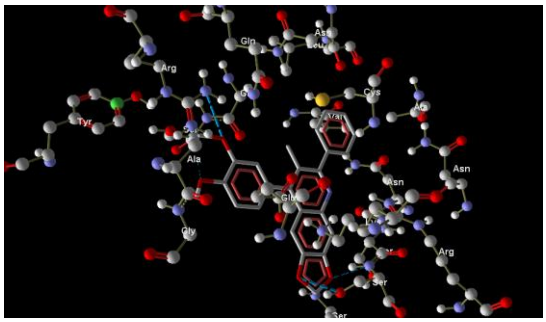


Figure 16

Number of Amino Acid : 345				
Molecular weight : 36230.1				
Theoretical PI : 4.89				
Amino Acids	Three letter Code	Single letter Code	No. of Amino Acids	Percentage of Amino Acids
Alanine	ALA	A	44	12.8%
Arginine	ARG	R	24	7.0%
Asparagines	ASN	N	9	2.6%
Aspartic Acid	ASP	D	23	6.7%
Cysteine	CYS	C	2	0.6%
Glutamic Acid	GLN	Q	12	3.5%
Glutamine	GLU	E	20	5.8%
Glycine	GLY	G	33	9.6%
Histidine	HIS	H	3	0.9%
Isoleucine	ILE	I	15	4.3%
Leucine	LEU	L	35	10.18%
Lysine	LYS	K	8	2.3%
Methionine	MET	M	5	1.4%
Phenylalanine	PHE	F	13	3.8%
Proline	PRO	P	21	6.1%
Serine	SER	S	24	7.0%
Threonine	THR	T	14	4.1%
Tryptophan	TRP	W	1	0.3%
Tyrosine	TYR	Y	3	0.9%
Valine	VAL	V	36	10.4%

Table 1

CARBON	C	1595
HYDDROGEN	H	2579
NITROGEN	N	453
OXYGEN	O	494
SULFUR	S	7

Table 2

Templates	Definition	No of amino acid residues	GeneID	Accession
3TZ6_A	Chain A, Crystal Structure Of Aspartate Semialdehyde Dehydrogenase Complexed With Inhibitor Smcs (Cys) And Phosphate From Mycobacterium Tuberculosis H37rv.	344aa	GI:388603970	3TZ6_A
3VOS_A	Chain A, Crystal Structure Of Aspartate Semialdehyde Dehydrogenase Complexed With Glycerol And Sulfate From Mycobacterium Tuberculosis H37rv.	362aa	GI:388604044	3VOS_A

Table 3

Sl no	Amino acid residues
1	ASP101
2	PRO102
3	VAL104
4	PRO105
5	LEU106
6	ASN112
7	ARG115
8	ASP116
9	ARG119
10	PRO121
11	LYS234

Table 4

Sl no	Sample code	BEFORE MINIMIZATION	AFTER MINIMIZATION
1	I-1	Energy=45.994164 Gradient=17.177937 Total energy=77.2804KCAL/MOL	Energy=25.376097 Gradient=0.095885
2	I-2	Energy=103.7918 Gradient=44.3105 Total energy=66.734KCAL/MOL	Energy=17.831448 Gradient=0.095419
3	I-3	Energy=111.849724 Gradient=29.913885 Total energy=81.5727KCAL/MOL	Energy=26.619295 Gradient=0.093283
4	I-4	Energy=100.556580 Gradient=44.054794 Total energy=72.3704KCAL/MOL	Energy=20.050508 Gradient=0.081373
5	I-5	Energy=121.554001 Gradient=56.320965 Total energy=69.1471KCAL/MOL	Energy=17.880527 Gradient=0.093808
6	I-6	Energy=122.605232 Gradient=56.034649 Total energy=68.4888KCAL/MOL	Energy=18.609116 Gradient=0.099888
7	I-7	Energy=101.444771 Gradient=43.987907 Total energy=70.0797KCAL/MOL	Energy=20.08228 Gradient=0.086729
8	I-8	Energy=176.03787	Energy=18.964077

		Gradient=90.530197 Total energy=73.2061KCAL/MOL	Gradient=0.088037
9	I-9	Energy=99.426010 Gradient=43.145699 Total energy=68.6695KCAL/MOL	Energy=19.582151 Gradient=0.098938
10	I-10	Energy=87.875465 Gradient=20.918631 Total energy=69.9302KCAL/MOL	Energy=18.471781 Gradient=0.099742

Table 5

Sl no	Sample code	Partial charges	Surface area (Approx)	Surface area (grid)	Volume	Hydration Energy	logp	Refractivity	Polarizability	Mass
1	I-1	0.00e	639.37	588.28	1001.66	-10.70	3.68	53.12	32.50A	366.27amu
2	I-2	0.00e	656.47	578.33	989.97	-5.32	4.38	56.05	32.18A	357.71amu
3	I-3	0.00e	652.84	624.36	1056.06	-10.60	4.49	54.22	34.88A	366.27amu
4	I-4	0.00e	647.15	588.55	1007.66	-5.36	4.81	52.57	32.55A	376.71amu
5	I-5	0.00e	669.72	604.85	1040.36	-5.45	5.419	57.16	34.57A	393.16amu
6	I-6	0.00e	668.72	600.31	1037.43	-5.31	5.19	57.16	34.57A	393.16amu
7	I-7	0.00e	647.73	587.18	1006.32	-5.27	4.81	52.57	32.55A	376.71amu
8	I-8	0.00e	677.96	612.52	1059.1	-5.48	5.46	59.98	35.27A	437..62amu
9	I-9	0.00e	616.85	600.65	1033.54	-5.36	4.29	51.77	32.25A	369.720amu
10	I-10	0.00e	589.59A	562.13A	984.55A	-5.92 KCAL/MOL	3.59	52.28A	30.51A	350.2amu

Table 6

Slno	Sample code	No of Hydrogen bond interactions	Amino acid connected to H bond	Energy of H bond	Length of H bond	Total energy of AA
1	I-1	1	LYS234	-0.693377	3.46132A	-27.3383
2	I-2	No interactions	-	-	-	-
3	I-3	4	Thr11 Gly12 Ser37 Ser37	-0.181176 -1.17106 -1.36793 -1.98239	3.39649 2.44003 2.46415 3.17525	0.645933 -9.37729 -11.5533 -11.5533
4	I-4	2	ARG147 LEU148	-1.5528 -2.37755	3.28944 3.12449	-14.03333 -17.9659
5	I-5	1	THR254	-0.912796	3.28383	-20.2592
6	I-6	3	ARG119 ARG119 ARG119	-0.0931469 -1.86122 -2.03809	3.57269 3.22776 3.14014	-5.39796
7	I-7	2	CYS130 SER71	-2.13639 -2.5	3.17272 3.06839	-4.90022 -9.07825
8	I-8	2	CYS130 ARG99	-2.4362 -1.26289	3.11348 3.31742	-7.1922 -5.24681
9	I-9	4	THR281 THR281 LYS140 VAL111	-2.5 -2.5 -0.4987 -1.7456	2.75276 3.0595 3.3414 3.2129	-13.6434 -6.80458 -9.54715
10	I-10	6	ARG249 ARG249 SER96 SE96 CYS130 SER71	-0.2937 -2.5 -2.5 -1.205 -2.13639 -2.5	3.08878 2.84436 2.75373 2.61211 3.17272 3.06839	-2.73845 -2.73845 -0.31231 -0.31231 -4.90022 -9.07825

Table 7

***Corresponding Author:**
E-mail: laksh.nov6@gmail.com