International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Structure Prediction of 3-phospho2-dehydro3-deoxyheptonate Aldolase from Mtb

Narendra Maddu[1] Anuradha Chevva Moremmagari[2]*

[1]Department of Biochemistry, Sri Krishnadevaraya University, Anantapur-515 003. AP, India.
[2]Department of Biotechnology, Sri Krishnadevaraya University, Anantapur-515 003. AP, India.

*Corresponding author e-mail:Anuradha.priyanka@gmail.com

Abstract

Comparative modelling involves prediction of 3D structures of large number of proteins. As traditionally defined, the term structural genomics is referred to the use of sequencing and mapping technologies, with the support of bioinformatics to develop complete genome maps and to elucidate genomic sequences of different organisms, particularly humans. Now, however, the term is increasingly used to refer to high thoroughput methods for determining protein structures. Pharmaceutical companies will be able to create drugs based on proteins, enzymes, and RNA molecules associated with genes and diseases. This will facilitate drug discovery and allow drug makers to produce a therapy more targeted to specific diseases. This accuracy not only will maximize therapeutic effects but also decrease damage to nearby healthy cells.

Keywords: Mycobacteriium, Aldolase, prediction, Drug.

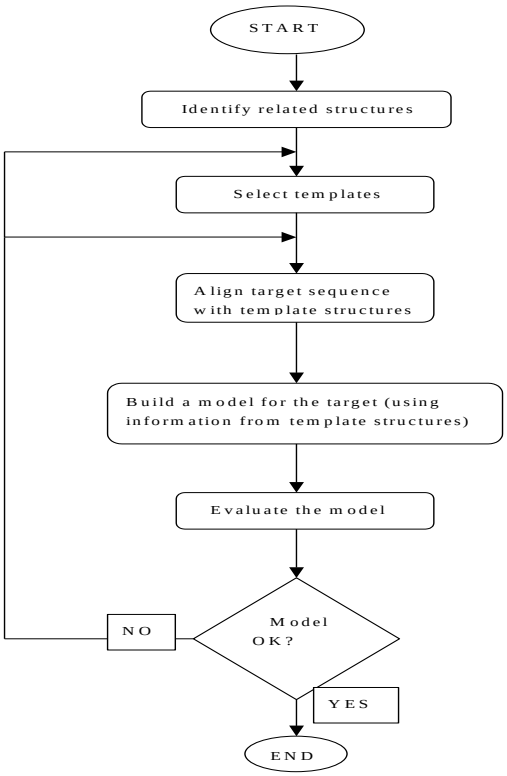
Introduction

Bioinformatics organizes data in a way that allows researchers to access existing information and to submit new entries as they are produced e.g. the Protein Data Bank for 3D macromolecular structures.While data curation is an essential task,the information stored in these databases is essentially useless until analysed thus the purpose of Bioinformatics extends far beyond mere volume control. Development of such resources requires extensive knowledge of computational theory,as well as thorough understanding of biology. Using these tools to analyse the data and interprete the results in a biologically meaningful manner and to conduct global analysis of all the available data with the aim of uncovering common principles that apply across many systems and highlight features that are unique to some. Instead of the standard trial-and-error method of matching patients with the right drugs, doctors will be able to analyze a patient's genetic profile and prescribe the best available drug therapy from the beginning. Not only will this take the guessworkout of finding the right drug, it will speed up recovery time and increase safety as the likelihood of adverse reactions is eliminated. The structures confirm the evolutionary changes in the primary structure of a given protein from related species, through random mutations. As these mutations lead to genetic disorder and diseases at the molecular level, clear understanding of the nature of these diseases is necessary. Also, when the structure of an enzyme is determined, a suitable inhibitor of the active site can be designed through combanitorial chemistry, computer modelling and docking techniques.This "structure based drug design" promises efficient drugs for several diseases in a short time. Evidences have established the fact that proteins undergo confirmational changes during their participation in the biochemical events. Highly resolved structures can lead to very clear understanding of the functions of these molecules.The structure-function relationship is key to our knowledge of the bio-world.

`

Given the large number of genes being discovered, the rate of new protein sequences is growing exponentially relative to the rate of protein structures being solved by experimental methods. Therefore alternative strategies like automated computational methods have to be employed in order to obtain 3-D structural information of the proteins. Comparative or homology modeling or knowledge-base prediction exploits the fact that evolutionarily related proteins with similar sequences have similar structures. While the high precision structures required for detailed studies of protein-ligand interaction can only be obtained experimentally, theoretically predicted models provide molecular biologists with " low resolution" models which hold enough information about preferred spatial arrangements of important residues to guide the design of experiments. Thus even though the current methods are still in their infancy, prediction of structures for all protein sequences of complete genomes in conjunction with experimental work is a realistic goal. Structural analyses of proteins for further mutagenesis, substrate and inhibitor design, and enhanced function and stability are also possible. These methods can use structural data to probe organism function and evolution.

Functional characterization of a protein sequence is one of the most frequent problems in biology. This task is usually facilitated by accurate three dimensional (3D) structure of the studied protein. There are two approaches to predict the structure of a protein. The comparative modeling procedure begins with an alignment of the target sequence with related known 3D structures. The output (obtained without any user intervention) is a 3D model for the target sequence containing all mainchain and sidechain non-hydrogen atoms. After a model is built, it is important to check it for possible errors. The quality of a model can be approximately predicted from the sequence similarity between the target and the template. Sequence identity above 30% is a relatively good predictor of the expected accuracy of a model.



The most successful protein structure prediction method to date is homology modeling and the approach is based on the structural conservation of the framework regions between the members of a protein family. Understanding the molecular function of proteins is greatly enhanced by insights gained from their three-dimensional structures. Since experimental structures are only available for a small fraction of proteins, computational methods for protein structure modeling play an increasingly important role. Comparative protein structure modeling is currently the most accurate method, yielding models suitable for a wide spectrum of applications, such as structure-guided drug development or virtual screening. Hence the preset work was carried out to predict the 3-D structure of 3-phospho2-dehydro3-deoxyheptonate aldolase of Mtb and to identify residues important for aldolase activity of Mtb which can be exploited in the design of more potent inhibitors.

TOOLS USED

BLAST : (http://www.ncbi.nlm.nih.gov/BLAST/)

BLAST (Basic Local Alignment Search Tool) is a set of similarity search programs designed to explore all of the available sequence databases. The scores assigned in a BLAST search have a well designed statistical interpretation. Making real matches easier to distinguish from random background hits. BLAST uses a heuristic algorithm, which seeks local as opposed to global alignments and is therefore able to detect relationships among sequences, which share only isolated regions of similarity.

PSI BLAST :

Position Specific Iterative BLAST (PSI-BLAST) refers to a feature of BLAST 2.0 in which a profile (or Position Specific Scoring Matrix PSSM) is constructed (automatically) from a multiple alignment of the highest scoring hits in an initial BLAST search.The PSSM is generated by calculating position-specific scores of each position in the alignment. Highly conserved positions recieve high scores and weakly conserved positions recieve scores near zero.The profile is used to perform a second BLAST search and the results of each 'iteration' used to refine he profile. This iterative searching strategy results in increased sensitivity

SWISS-PROT : (http://www.expasy.org/sprot/)

SWISS-PROT is an annotated protein sequence database with protein knowledgebase consists of sequence entries. sequence entries are composed of different line types, each with their own format. For standardization purposes the format of SWISS-PROT follows as closely as possible that of the EMBL nucleotide sequence database.

PDB : (http://www.rcsb.org)

It is the international repository for the processing and distribution of 3D macromolecular structure data primarily determined by X-ray crystallography and NMR. The PDB archive contains macromolecular structure data on proteins, nucleic acids, protein-nucleic acid complexes, and viruses. A variety of information associated with each structure is available, including sequence details, atomic coordinates, crystallization conditions, 3D structure neighbors computed using various methods, derived geometric data, structure factors, 3D images, and a variety of links to other resources.

PROSITE : (http://www.expasy.org/PROSITE/)

PROSITE is an ExPASy database that catalogs biologically significant sites through the use of motif and sequence patterns known as fingerprints. This database is used in profile analysis. PROSITE is a method for determining the function of uncharacterized proteins translated from genomic or cDNA sequences. This comprises of a database of biologically significant sites and patterns formulated in a manner that

with appropriate computational tools it can rapidly and reliably identify to which known family of protein (if any) the new sequence belongs.

Pfam : (http://www.sanger.ac.uk/cgi-bin-/Pfam/pfamb ast_server). It is a manually curated collection of protein families available via the web and in flat file form. Genome projects including both the human and fly, have used Pfam extensively for large scale functional annotation of genomic data.

MODELLER^R :

MODELLER^R is a computer program that models 3D structure of proteins by satisfaction of spatial restraints. MODELLER^R is most frequently used for homology or comparative protein structure modeling. The user provides an alignment of a sequence to be modeled with known related structures and MODELLER^R will automatically calculate a model with all non-hydrogen atoms. Generally, the input to the program are restraints on the spatial structure of the amino acid sequence(s) and ligands to be modeled. The output is a 3D structure that satisfies these restraints as well as possible. Restraints can in principle can be derived from a number of different sources. These include related protein structures (comparative modeling), NMR experiments, cross-linking experiments, fluorescence spectroscopy, image reconstruction in electron microscopy, site directed mutagenesis, intuition, residue-residue and atom-atom potentials of mean force. The restraints can operate on distances, angles, dihedral angles, pairs of dihedral angles and some other spatial features defined by atoms or pseudo atoms. Presently, MODELLER automatically derives the restraints only from the known related structures and their alignment with the target sequence. A 3D model is obtained by optimization of a molecular probability density function (pdf). Fig. 4).The molecular pdf for comparative modeling is optimized with the variable target function procedure in cartesian space that employs methods of conjugate gradients and molecular dynamics with simulated annealing. MODELLER can also perform multiple comparison of protein sequences and/or structures, clustering of proteins, and searching of sequence databases.

PROCHECK

It checks the stereo-chemical quality of the structure . It is based on an analysis of angles using Ramchandran map (Ramchandran and Sasisekharan, 1968), peptide bond planarity, bond lengths, bond angles, hydrogen-bond geometry, and side chain conformations of known protein structures as a function of atomic resolution.

Procheck:

The Procheck is obtained from the web site www.rcsb.org from validation server.

Ramachandran plot :

Ramachandran plot is obtained from the web site www.rcsb.org from validation server.

Verify 3D:

The modeled structure was submitted to VERIFY_3D a visual analysis of the quality of a putative crystal structure for a protein was provided in fig .6

All the values from procheck and Ramachandran plot values were compared with the X-ray crystallography 3D structure.

RESULTS

BLAST is done to obtain the template sequence. As the target sequence is having no structure the template sequence having more homology is selected and homology modeling of the template sequence is done. The blast results are shown below in the fig.2

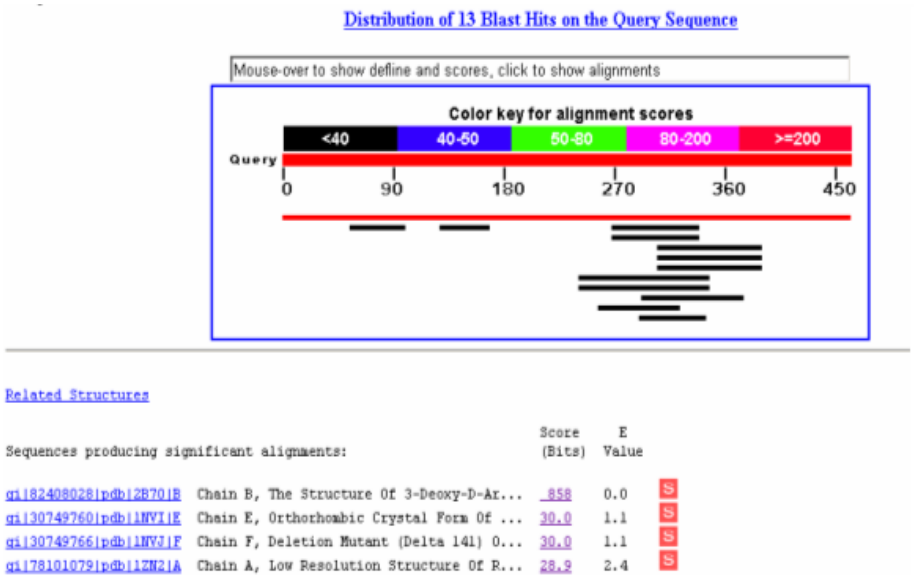


Fig.2. The blast results of Aldolase of Mtb

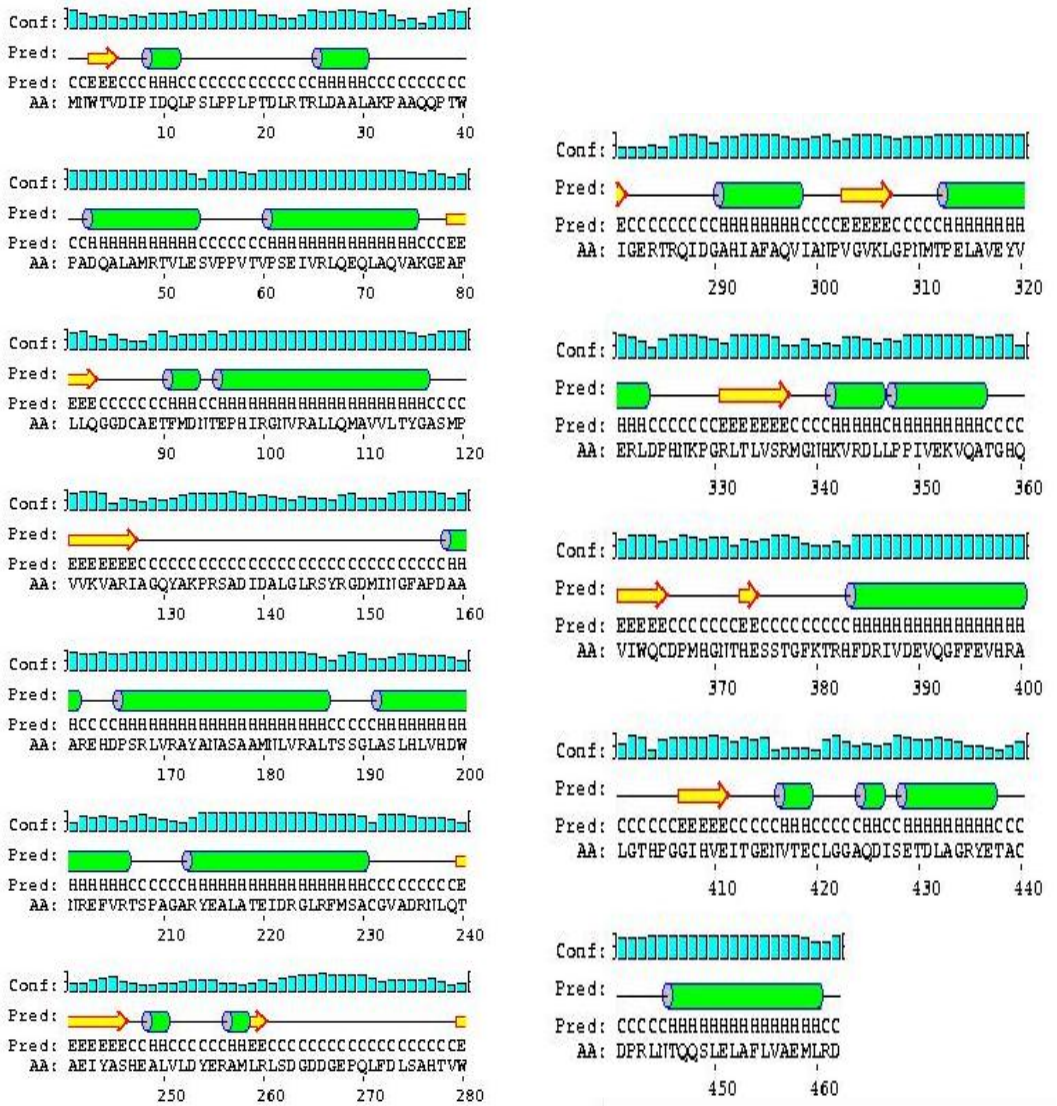


Fig.3. Secondary Structure prediction of Aldolase of Mtb

ScanProsite Results Viewer

This view shows ScanProsite results together with ProRule-based predicted intra-domain features (help).

show hits of frequently occuring signatures

Hits for all PROSITE (release 19.26) motifs on sequence aldolase :

found: 17 hits in 1 sequence

ALDOLASE (462 aa)

```
MNWTVDIPIDQLPSLPPLPTDLRTRLDAALAKPAAQQPTWPADQALAMRTVLESVPPVTVPSEIVR
LQEQLAQVAKGEAFLLQGGDCAETFMDNTEPHIRGNVRALLQMAVVLTYGASMPVVKVARIAGQYA
KPRSADIDALGLRSYRGDMINGFAPDAAAREHDPSRLVRAYANASAAMNLVRALTSSGLASLHLVH
DWNREFVRTSPAGARYEALATEIDRGLRFMSACGVADRNLQTAEIYASHEALVLDYERAMLRLSDG
DDGEPQLFDLSAHTVWIGERTRQIDGAHIAFAQVIANPVGVKLGPNMTPELAVEYVERLDPHNKPG
RLTLVSRMGNHKVRDLLPPIVEKVQATGHQVIWQCDPMHGNTHESSTGFKTRHFDRIVDEVQGFFE
VHRALGTHPGGIHVEITGENVTECLGGAQDISETDLAGRYETACDPRLNTQQSLELAFLVAEMLRD
```

ruler: 1 100 200 300 400 500 600 700 800 900 10

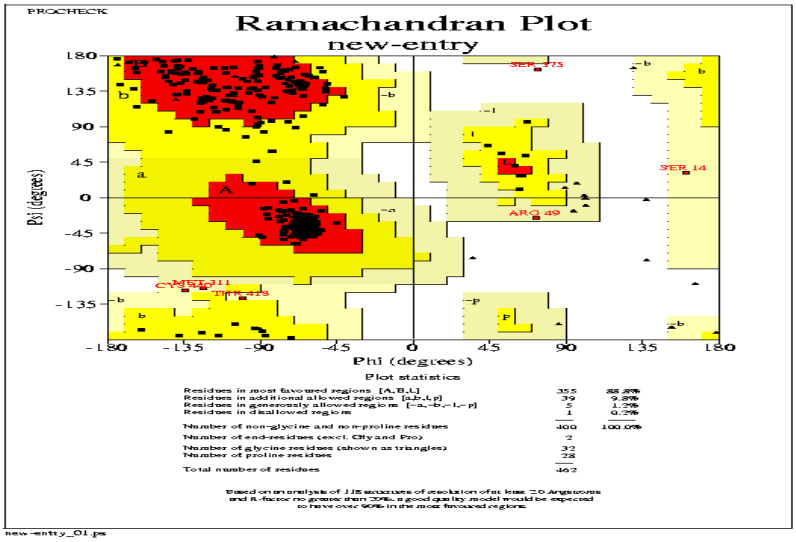
hits by patterns with a high probability of occurrence or by user-defined patterns: [17 hits (by 6 distinct patterns) on 1 sequence]

ALDOLASE (462 aa)

PROCHECK and RAMACHANDRAN plot analysis

The Ramachandran plot is obtained from www.rcsb.org.

The target sequence is subjected to validation server to obtain the Ramachandran plot. The residues in the most favored regions are 88.8%. The residues in the additional allowed regions are 9.8%. The residues in generously allowed regions are 1.2%. The residues in disallowed regions are 0.2%. 88.8% bond lengths are with in limits. 72%bond angles are with in limits.



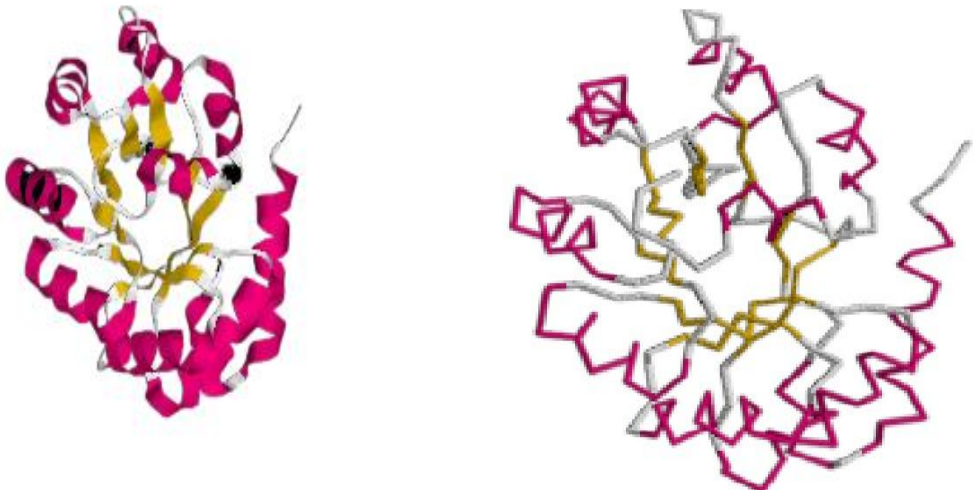


Fig. 4. Predicted 3-D Structure of 3-phospho2-dehydro3-deoxyheptonate aldolase
(A) Ribbon display, (B) Chain display.

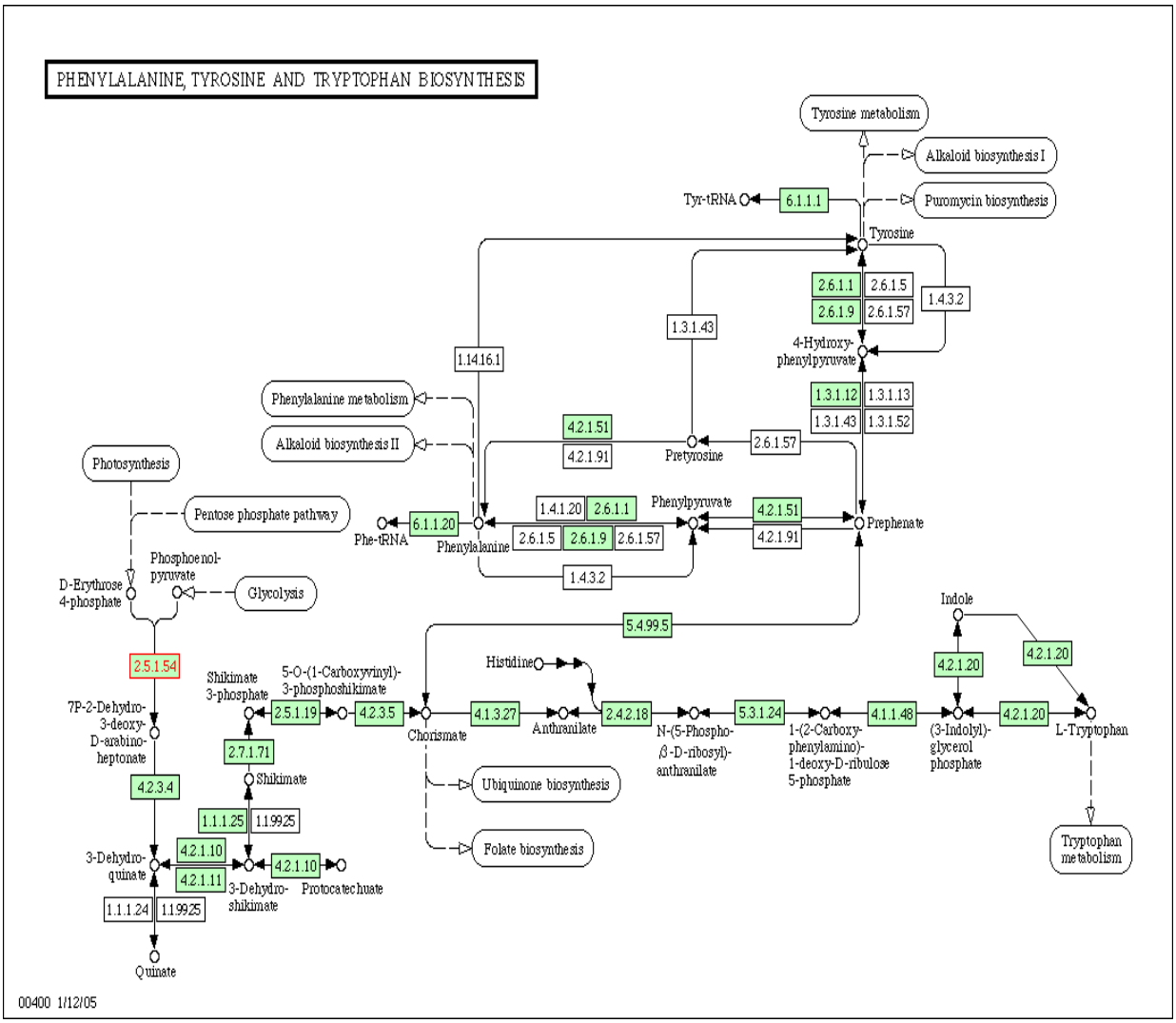
Discussion:

Tuberculosis is caused by *Mycobacterium tuberculosis*, a leading cause of mortality due to a bacterial pathogen. The shikimate pathway is an attractive target for the development of antimycobacterial agents, because it has been shown to be essential for the viability of mycobacterium tuberculosis ,but absent from mammals. In myco bacteria, the shikimate pathway leads to the biosynthesis of chorismic acid which is a precursor of aromatic amino acids, napthoquinone, menaquinones and mycobactins.

The shikimate pathway comprises a series of seven enzyme catalysed reactions that result in the biosynthesis of chorismate which is the precursor for many essential aromatic amino acids tyrosine tryptophan phenylalanine folic acid an essential cofactor for many enzymatic processes and salicylate used for the biosynthesis of the siderophores through which bacteria acquire iron. The first committed step in the shikimate pathway is catalyzed by 3-phospho2-dehydro3-deoxyheptonate aldolase. The 3-phospho2-dehydro3-deoxyheptonate aldolase catalysis a stereo specific aldol like condensation between phospho enol pyruvate and erythrose 4-phosphate to give DAH7P3 and inorganic phosphate.These are revealed through labeling studies and through structural analysis of the DAH7P synthases from *E.coli*

Analysis of the genome of *M. tuberculosis* reveals the presence of a single open reading frame encoding a putating DAH7PSs of 462 residues. This belongs to the type II DAH7PSs.This type II DAH7PSs are required for the aromatic amino acid biosynthesis in mycobacterium tuberculosis bacteria. This enzyme might have been catalyzing the similar step of shikimate pathway that occurring in the *E.coli.*

This enzyme is found to be metal ion dependent and subjected to feedback inhibition by phenyl alanine ,tyrosine,tryptophan and chorismate. Glyphosate is a successful, broad spectrum, postemergence herbicide. It is believed to disrupt aromatic amino acid biosynthesis in plants by reducing metabolic flux in the shikimate pathway. The DHAPand DHQ synthases present in bacterial extracts can be inhibited by concentrations of glyphosate in the mM range.



Homology modeling is a multistep process which converts a linear amino acid sequence of a protein into a 3 dimensional structure. The amino acid sequence of 3-phospho2-dehydro3-deoxyheptonate aldolase (target) was taken from GenBank (NCBI) and subjected to similarity search through BLAST group of programs to identify the sequences with highest homology. The most homologus sequence is the one with lowest E- value among all the hits.This sequence was selected as the template. The FASTA format of both the target and template was loaded on CLUSTALX software to obtain the alignment.

The secondary structure prediction results (Fig. 2) were obtained for the aldolase of Mtb and depicted the helical structures more in nature and number. The 3-D structure of 3-phospho2-dehydro3-deoxyheptonate aldolase of Mtb (Fig. 4) showed the presence of alpha-helices, beta-pleated loops (Fig. 3). The predicted 3-phospho2-dehydro3-deoxyheptonate aldolase 3-D structure was then evaluated for possible errors by PROCHECK and Ramachandranplot (Fig. 4). The structure was satisfactory as per the Ramchandran plot constraints. The final built model of 3-phospho2-dehydro3-deoxyheptonate aldolase of Mtb shown in Fig. 5. It satisfied the G value as well as other constraint values. Mt is difficult to kill for a number of reasons such as its slow growth, dormancy, complex cell envelope, intracellular pathogenesis and genetic homogeneity [1]. Very few drugs have been developed in the past 44 years [2, 3]. There is thus a profound need for the identification and development of novel chemotherapeutic compounds active

against TB. The availability of complete genome sequence of Mt provides a better understanding of the biology of this slow pathogen and also unravels potential targets that may be of utility in prophylactic and therapeutic interventions. This will facilitate drug discovery and allow drug makers to produce a therapy more targeted to specific diseases. This accuracy not only will maximize therapeutic effects but also decrease damage to nearby healthy cells. Instead of the standard trial-and-error method of matching patients with the right drugs, doctors will be able to analyze a patient's genetic profile and prescribe the best available drug therapy from the beginning. Not only will this take the guessworkout of finding the right drug, it will speed up recovery time and increase safety as the likelihood of adverse reactions is eliminated. Selective and specific Mtb-MAT inhibitors can be designed using the homology model, by the structure based drug design approaches.

References:

1. Altschul SF et al., J Mol Biol, 2019. 215: 403-410.
2. Altschul SF et al., Nucleic Acids Res, 2019. 25: 3389-3402.
3. Bentley andCrit Rev. ***Biochem Mol. Biol*** 2018 25, 307-384.
4. Bersa G.S. et al., ASM *Press, Washington*, 2019, pp. 285-306.
5. Butler WR, Guthertz LS, *Clin Microbiol Rev* 2018, **14:**704-26.
6. Coggins et al., ***Biochem. Soc Trans*** 2018, 31, 548.
7. Cole S.T et al., *Nature* 2017, **409:**1007-11.
8. Corbet.E.L et al., *Arch. Int. Med*. 163 (2018) 1009-1021.
9. C.Ratledge and Stanford, ***The Biology of the Mycobacteria, Vol.1,Academic Press, London***, 2019 , pp 185-271.
10. Fraser C.M et al., *Science* 2019, **270:**397-403
11. JJ.de Voos et al., ***Natl. Acad.Sci USA*** 97 2017, 1252-1957.
12. Jin, D.J., and C. A Gross. 2018. J. Bacteriol.171: 5229-5231.
13. Jones G et al., J Mol Biol 2019; 267:727–748.
14. Konig V et al**.,** ***Mol.Biol***. 2004, 337,675.
15. Kotb.M et al., *Pharmacol. Ther*. 59 (1993) 125-143.
16. Mato.J.M et al., Pharmacol. Ther. 73 (1997) 265-280.
17. Marton.L.J et al., *Ann. Rev. Pharmacol. Toxicol*. 2019. 5: 55-91.
18. Needleman SB et al., J Mol Biol, 2018. 48: 443–453
19. Laskowski.R.A RA, MacArthur MW, Moss DS & Thorton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26. (2019); 283-291.
20. Lüthy R, Bowie JU, Eisenberg D, Nature, 2019. 356 ; 83-85.
21. Parrish NM, Dick JD, Bishai WR, *Tr Microbiol* 2019, **6:**107-12
22. Philip.W.J et al., Proc. Natl. Acad. Sci. U.S.A. 2018. 93: 3132-3137
23. Raviglione, The TB epidemic from 1992 to 2020, tuberculosis 83 2003, 4-14.
24. Sarkar.N.K et al., *Biochem.Mol.Bio.Int*. 2018. 35 : 1189-1198.
25. Sali & T. L. Blundell, J. Mol. Biol., 2016. 234: 779.
26. Schofield et al., ***Biochemistry*** 2015, 44, 11950.
27. Shflyan et al., ***Chem. Soc***. 2019,120,11027.
28. Shumilin et al., [London] 2019, 7, 865.

29. Shumilin et al., ***Mol.Biol***. 2018,320, 1147.
30. Shumilin et al., ***Biochemistry*** 2016, 42, 3766.
31. Shumilin et al., ***Mol.Biol***. 2019,341,455.
32. Shigenobu .S et al., *Nature* 2017, **407:**81-6.
33. Srinivasan P.R.and Sprinson,2-keto-3-deoxy-D-arabo-heptonic acid 7-phosphate synthase**,** Steinrucken and Amrhein 2018, ***Biochem. Biophys. Res.Comman***. 94, 1207-1212.
34. SYBYL 6.7, Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA
35. Tabor.C.W et al., *Adv. Enzymol*. 2019. 56 : 251-282.
36. Takusagawa.F et al., Biochemistry. 2018. 35 :2586-2596.
37. Takugasawa.F et al., J. Biol. Chem. **2017**. 271: 136-147
38. Thompson JD et al., Nucleic Acids Res, 25: 2019; 4673-4680.
39. Toscano et al., ***C.org.Biomol.Chem***. 2018,I, 2075
40. Webby et al., ***Mol. Biol*** 2017, 353, 927.
41. Williamson et al., Bioorg. ***Med.chem. Lett*** 2015, 15, 2339.
42. WHO, Tuberculosis Fact Sheet (No. 104), 2019.