ISSN 0975 - 6299 Vol 2 / Issue 4 / Oct - Dec 2011

International Journal of Pharma and Bio Sciences

RESEARCH ARTICLE BIOINFORMATICS

COMPARATIVE GENOMICS OF B- SUBUNIT OF *TRYPTOPHAN SYNTHASE* AMONG A GROUP OF PROTEOBACTERIA

SATPAL SINGH BISHT*, SANTOSH KUMAR BEHERA, AMRITA PANDA AND B.PRAVEEN.

Department of Biotechnology, Roland Institute of Pharmaceutical Sciences Berhampur-760010, Orissa, India

SATPAL SINGH BISHT
Department of Biotechnology, Roland Institute of Pharmaceutical Sciences Berhampur-760010, Orissa, India

ABSTRACT

Proteins are macromolecules with a characteristic sequence of amino acid linked by peptide bond and play a vital role in metabolic pathways. There are 20 major amino acids and hundreds of "minor amino acids" - some are short-lived and some are the modifications of the main 20. Tryptophan synthase, in the biosynthetic pathway of the aromatic amino acid Tryptophan is regulated by the “trp operon” in E.coli. Tryptophan: In the *Shikimic acid pathway*, Chorismate is a key intermediate in the biosynthesis leading to the formation of the aromatic amino acids (*phenylalanine, tyrosine and tryptophan*). *Tryptophan Synthase* is a bifunctional, bienzyme complex having two α chains and two β chains that forms the $\alpha_2\beta_2$ tetrameric enzyme complex. In most of the organisms α chain is encoded by the gene trpA and β chain is encoded by the gene trpB. *Salmonella typhimurium* was taken as a reference organism for comparative genomics of Tryptophan synthase and data that was available for comparative genomics was interrogated with bioinformatics tools. The amino acid (aa) sequence of tryptophan synthase beta chain of *Salmonella typhimurium* from the PDB was taken as the standard sequence and pair-wise alignments and multiple sequence alignment was pursued for all the 41 proteobacteria using BLAST 2 programme and Clustal X (1.83). phylogenitic analysis was drawn by using tree view.

ISSN 0975 - 6299 Vol 2 / Issue 4 / Oct - Dec 2011

KEY WORDS

Salmonella typhimurium, Proteobacteria, PDB (Protein Data Bank), *Tryptophan synthase*, Bioinformatics

INTRODUCTION

In the *Shikimic acid pathway*, Chorismate is a key intermediate in the biosynthesis leading to the formation of the aromatic amino acids (*phenylalanine, tyrosine and tryptophan*). The metabolic pathways of the three aromatic amino acids branch out from chorismate therefore chorismate is often considered a limiting factor in the formation of tryptophan, with much of the compound being utilized for a variety of other metabolites.

Tryptophan is synthesized in bacteria, *E.coli,* by a five-step biochemical pathway from Chorismate. Five enzymes are involved in the biosynthesis of tryptophan and there are five genes in the tryptophan operon in *E.coli* that code for polypeptide chains which organises themselves to make these five enzymes. Tryptophan synthase catalyzes the last step of the Tryptophan biosynthetic pathway that involves conversion of Indole-3-glycerol phosphate and Serine to Tryptophan and water [1]. It is a bifunctional, bienzyme complex having two α chains and two β chains that forms the $\alpha_2\beta_2$ tetrameric enzyme complex. In most of the organisms α chain is encoded by the gene trpA and β chain is encoded by the gene trpB.

The crystal structure of the Tryptophan Synthase explained by C. Craig Hyde et.al [2]. The sequence and three-dimensional structural information of the enzyme Tryptophan synthase is for the proteobacteria *Salmonella typhimurium* (Fig1). The 3D structure of $\alpha_2\beta_2$ complex of tryptophan synthase from *Salmonella typhimurium* has been determined by X–ray crystallography at 2.5 a resolution. The four polypeptides are arranged linearly in $\alpha\beta\beta\alpha$ order forming a complex 150A° long .The two tightly associated β subunits are at the center of the complex, while the two α subunits are located distant from each other at the opposite ends of the β_2 dimer .The length of the α subunit is 268 amino acid residues, where as the length of the β subunit is 397 amino acid residues. The total length of the αβ dimer is 665 amino acid residues. The α subunits are smaller then β subunits. The active site of each α subunit is located near the interface with the subunit, where as the active site of the β subunit is deeply buried in the center of the β subunit. The two active sites of neighboring α and β subunits are separated by a distance of about 25-30 A° .The α subunit which has an 8 fold α/β barrel fold is *tilted* relative to the β subunit such that the central axis of the α subunit barrel forms an appropriate 20° angle with a line drawn between the two active sites. The active sites of the two-β subunits in the $\alpha_2\beta_2$ complex are quite distant (about 25°) and it seems unlikely that residues from one β subunit contribute to the active site of the other β subunit. The two-α/β subunit pairs in a tetramer are in a sense back to back with each other. Thus, the formation of $\alpha_2\beta_2$ tetramer does not appear necessary for catalysis but may play a role in stabilizing the conformation of the β subunits. The β subunit has two domains - N domain and C domain. Part of the N-domain of each β

subunit interacts with part of the C domain of the complementary β subunit. The β subunit usually exists as a dimer, contains two molecules of the cofactor Pyridoxal phosphate (PLP) per dimer, and catalyzes the synthesis of L-tryptophan from indole and L- serine, termed the β reaction. When the α and β subunits combine to form the $\alpha_2\beta_2$ complex, the rates of α and β reactions and the affinities for substrates in these reactions are increased by 1-2 orders of magnitude .The physiologically catalyzed by the $\alpha_2\beta_2$ complex is termed the α-β reaction, is the sum of the α and β reactions.

The interface between α and β subunits is mostly hydrophobic in character. There is a 25 A° long hydrophobic tunnel through which the metabolic intermediate of αβ-reaction, indole, is transferred from α to β -subunit .The tunnel not only acts to direct the diffusion of indole in order to prevent its loss by diffusion through cell membrane, but also takes part in allosteric interactions that synchronize the reactions taking place at both the ends of the tunnel. Tryptophan Synthase is a pyridoxal dependent enzyme, so the reaction of the Serine with the co-factor pyridoxal phosphate at the β-site modulates the formation of indole at the α-site, as well as its passage through the tunnel, so that indole does not reach the β-site before a highly reactive aminoacrylate has been formed .

The co-enzyme pyridoxal phosphate covalently binds to the β lys87 and forms schiff base but when serine binds, the amino group of serine replaces the ε-amino group of lysine forming an external aldimine [3]. This is followed by a deprotanation step, which leads to a quinoid that is subsequently protanated by loosing water. This leads to formation of an aminoacrylate. If the highly reactive aminoacrylate gets an indole, it produces Tryptophan and internal aldimine. In the absence of indole, the aminoacrylate slowly hydrolyzes to pyruvate, ammonia and internal aldimine.

MATERIALS AND METHODS

The online tools used are BLAST 2- (Basic Local Alignment Search Tool.), ClustalX (1.83) RasMol, Tree Viewer, RasTop, Ghost View The information related to the tryptophan biosynthesis pathway in a number of organisms is available in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. Most genomes and information are for prokaryotes (proteobacteria) and few are of archea and eukaryotes. Forty one proteobacteria were selected for analysis of enzyme Tryptophan synthase present in most of these organisms the alpha subunit of tryptophan synthase was encoded by the gene trpA and the beta subunit is encoded by the gene trpB. In some organisms, like *Coxiella burnetii* (cbu), the trpB gene is fused with phosphoribosyl anthranilate isomerase gene (trpF). In the other taxonomic groups certain organisms like *Pyrococcus furiosus* (Pfu) and *Vibrio parahaemolyticus* (vpa) in archea two different genes encode for the two beta chains of tryptophan synthase, and in *Saccharomyces cerevisiae* (yeast) the α and β subunits are encoded by a single gene trp5.

Amino acid sequence beta subunit of tryptophan synthase from *S. typhimurium* was taken as the standard sequence from the Protein Data Bank (PDB) as its crystal structure is known. Performed pair-wise sequence alignment using BLAST 2 and multiple sequence alignment of the amino acid sequences using ClustalX for all the 41 proteobacteria with respect to the amino acid sequence of *S. typhimurium* to extract information about the function of the enzyme

tryptophan synthase in different proteobacteria[4].

Multiple sequence alignment was done by ClustalX (1.83) to find out the conserved residues[5]. Amino acid sequences of the protein tryptophan synthase beta subunit of 41 proteobacteria taken from KEGG[6] followed by the amino acid sequence of protein tryptophan synthase beta subunit of *Salmonella typhimurium* from PDB[7] (code 1BKS supersedes 1WSY) as the standard sequence for counting the positions of the residues in multiple alignment work. The ClustalX programme provides an integrated environment for performing multiple sequence alignments and analyzing the results. Clustal X provides a versatile coloring scheme for the display of the alligned sequences.

RESULTS AND DISCUSSION

The percentage of identity ranged from 47% - 94%. The gaps inserted to get the significant alignment were always less then 3%. In some alignments, with longer and shorter sequences, the positions of the specific residues were changed because of difference in the sequence length e.g in the alignment of stm with cje, the residue glutamine at 205 in stm was aligned with glutamine at 202 in cje. Similarly, in the alignment of stm with mlo, the residue phenylalanine at 6 in stm was aligned with phenylalanine at 20 in mlo. The results obtained from multiple sequence alignment are shown. The analysis of the result shows that there is considerable amount of conservation seen in the amino acid residues in proteobacteria. The findings revealed that **198** totally conserved positions, which are represented in the form of ., ~ and (:) The star represents completely conserved positions, dot represents mostly conservative residues and the colon represents acceptable substitutions respectively of the **198** conserved residues, there are **97** completely conserved positions, **74** acceptable substitutions and **27** mostly conserved and there are many conserved positions except in few organisms, for example, positions **378G, 379R, 380G** and **382K** are completely conserved except in one organism i.e. *Buchnera aphidicola* (Buc).

In order to find the evolutionary relationships among these 41 proteobacteria with respect to amino acid sequence of beta chain of tryptophan synthase, cladograms constructed with the help of tree viewer tool and ClustalX. The cladograms are the branching, treelike diagram in which the endpoints of the branches represent specific species of organisms. It is used to illustrate phylogenetic relationships and show points at which various species have diverged from common ancestral forms.

The cladogram (Fig-2) shows that *S.typhimurium*, *S. typhi_Ty2, S. typhi* are located on a single branch and indicates the close relation among these organisms with respect to amino acid sequence of beta chain. The organism *Campylobacter jejuni* (cje) is located in the other branch, which is far distant from *S.typhimurium*, this states how distantly both the organisms are related with respect to the amino acid sequence of beta chain of tryptophan synthase.

Sequence analysis results of beta chain from 41 proteobacteria with information obtained from the structural analysis of beta chain of tryptophan synthase in *S. Typhimurium* and inferred about the possible functional correspondence of the conserved residues.

- The residues **Gly232, Gly233, Gly234, Ser 235&Ala237** and **Asn 236** (between strand 7 and helix 9) are conserved**.** The phosphate group of the co-enzyme is highly ligated through hydrogen bonds with the peptide backbone atoms of these residues.

- The residue **His 86** (before N terminal helix3) is found to be conserved in all the organisms**.** The negative charges on phosphate may be neutralized by the imidazole of **His 86** and the positive end of a dipole from helix 9.
- The residue **Lys 87 (**at the N-termini of helix3) is also found to be conserved in all the organisms. The co-enzyme PLP forms a covalent bond with ε-amino group of **Lys 87** of β subunit.
- The residues **267His, 269Ala, 274Gly, 277Gly, 281Gly, 292Gly, 299Ser, 303Gly, 304Leu, 305Asp, 307Pro and 310Gly** (between strand 8 and helix 10) are conserved in all the organisms. These residues fold in a complicated way and apparently lack any well defined secondary structural elements and these residues are involved formation of wall of the tunnel. The residues **186Y, 189G, 190T** (between helix 6 and strand 6) are conserved as they are involved in formation of wall of the tunnel.
- The residues **86His, 87Lys, 89Asn, 97Leu, 98Ala, 101Asn, 103Lys, 107Ile 108Ala, 109Glu, 110Thr, 111Gly, 112Ala, 113Gly, 115His, 116Gly, 118Ala, 120Ala, 124Ala, 130Cys, 135Gly** (residues from helix 3,strand 3,helix 4and strand 4) and **145Asn, 146Val, 149Met, 153Gly, 154Ala, 162Gly, 166Leu, 167Lys, 169Ala, 173Ala, 186Tyr, 189Gly** (residues from helix 5, strand 5, helix 6 and strand 6) are found to be structurally similar with the residues **204Phe, 209Gly, 211Glu, 215Gln, 229Ala, 232Gly, 233Gly, 234Gly, 235Ser, 236Asn, 237Ala, 239Gly, 241Phe, 250Val, 254Gly, 256Glu, 259Gly** (residues including helix 8 to strand 8 along with strand 7,helix 9) and **307pro, 310Gly, 311Pro, 316Lys, 321Arg, 324Tyr, 329Asp, 331Glu, 335Ala, 343Glu, 344Gly, 345Ile, 348Ala, 350Glu, 353His, 354Ala, 356Ala** (residues including helix 10 to helix 12 along with strand 9 ,helix 11) so these residues are found to be conserved in all the organisms.

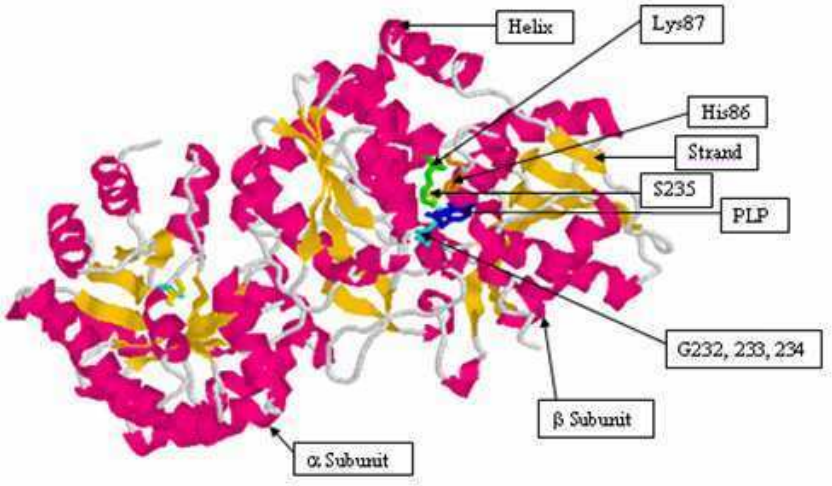

Fig.1

Schematic view of alpha and beta subunits of tryptophan synthase in Rasmol in cartoon format.

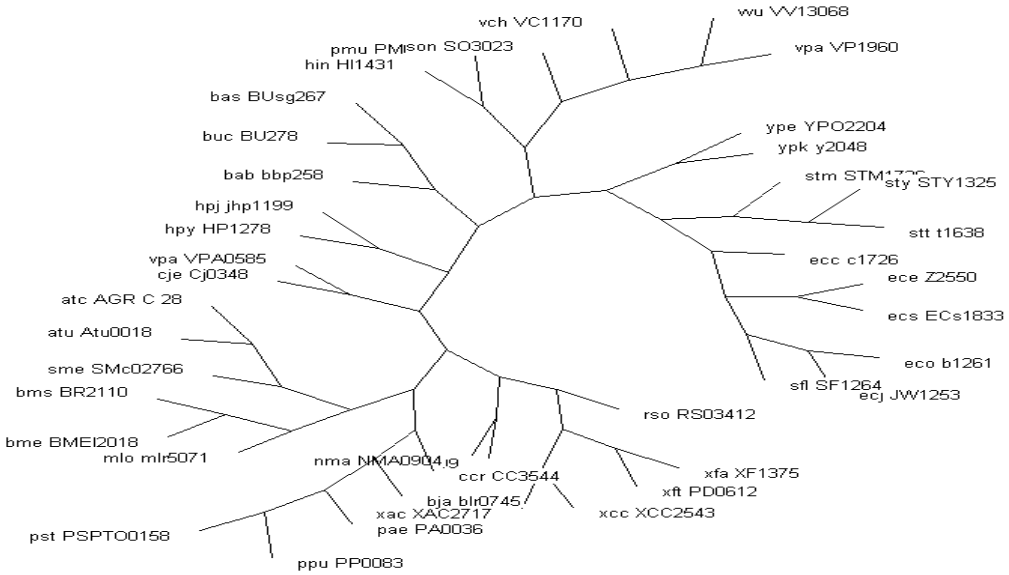

Fig. 2
Radial tree (un-rooted tree) view of 41 proteobacteria

ACKNOWLEDGEMENTS

The authors are very much grateful to Roland Institute of Pharmaceutical Sciences for providing necessary research facilities.

REFERENCES

1. Nelson David, L. and Michel M.Cox, Lehninger Principles of Biochemistry,3rd edn, Worth Publishers:836,(2002)
2. Hyde Craig C., Ashrafudin Ahmed S., Eduardo A. Padlan, Miles Edith, Wand Davies David R. Three-dimensional Structure of the Tryptophan Synthase $\alpha_2\beta_2$ Multienzyme Complex from *Salmonella typhimurium*, The Journal of Biochemistry,263: 17857-17871, (1988).
3. Schneider Thomas R., Gerhardt E., Lee M., Liang P.H, Anderson K. S and Schlichting I. Loop Closure and Inter-subunit Communication in Tryptophan synthase, Biochemistry, 37: 5394-5406,(1998).
4. Mount David W. *Bioinformatics*: Sequence and Genome Analysis, CSHL, New York: 75-85,156,(2001).
5. Westhead D.R., Parish J.H.,Twyman R.M. Instant Notes on Bioinformatics,1st edn,BIOS Scientific Publishers Limited: 83-85,(2003).
6. http://www.genome.ad.jp/kegg/pathway.html
7. http://www.rcsb.org/pdb/