

## Homology modeling of phosphoryl thymidine kinase of enterohemorrhagic *Escherichia coli* O157:H7

Shilpa Deshpande Kaistha<sup>1,2,\*</sup> and Ranjana Sinha<sup>1</sup>

<sup>1</sup>Department of Bioinformatics, University Institute of Engineering Technology; <sup>2</sup>Department of Microbiology, Institute of Biosciences and Biotechnology, Chhatrapati Sahuji Maharaj University, Kanpur 208024, UP, India; Shilpa Deshpande Kaistha\* - Email: shilpakaistha@gmail.com; \* Corresponding Author

received September 17, 2008; revised November 11, 2008; accepted December 01, 2008; published January 12, 2009

### Abstract:

Enterohemorrhagic *Escherichia coli* (EHEC) are source of emerging infectious disease in India. *Escherichia coli* O157:H7 is an EHEC strain showing multiple antibiotic resistances and the cause of infantile diarrhea and hemolytic uremic syndrome worldwide. A novel strategy to counteract multiple antibiotic resistant organisms is to design drugs which specifically target metabolic pathways such as thiamine biosynthetic pathways found exclusively in prokaryotes. Homology modeling was used for model building of a terminal thiamine biosynthesis enzyme phosphoryl thymidine kinase (Thi E) using Geno3D, Swiss Model and Modeller. The best model was selected based on overall stereochemical quality. The potential ligand binding sites in the model were identified by CASTp server. The validated theoretical model of the 3D structure of the thiE protein of *E. coli* O157:H7 was predicted using a thiamine phosphate pyrophosphatase from *Pyrococcus furiosus* (PDB ID: 1X13\_A) as template. The active pockets of ligand binding sites in the enzyme were identified. In this study, phosphoryl thymidine kinase (thi E), a terminal enzyme in the thiamine biosynthesis pathway in the pathogen has been modeled to be used in future as a potential drug target by the design of suitable inhibitors.

**Keywords:** EHEC; thi E; phosphoryl thymidine kinase; homology modeling

### Background:

Enterohemorrhagic *Escherichia coli* (EHEC) are a group of Gram Negative bacteria belonging to Enterobacteriaceae family that are a significant cause of infantile diarrhea and hemolytic uremic syndrome in the developing and developed nations [1]. Infections are usually transmitted following ingestion of unhygienic packed and processed food, milk and water resources. One typical EHEC, *Escherichia coli* O157:H7, also known as verotoxigenic *Escherichia coli* (*E. coli*) or shiga like toxin producing *E. coli*, is the causative agent of hemorrhagic colitis [1]. These organisms have been characterized to produce virulence factors like periplasmic catalases and shiga-like toxins which act by catalytically inactivating 60S ribosomal subunit of the eukaryotic protein translation machinery [1, 2]. The verotoxin itself is expressed by a prophage carried by the *E. coli* strains. The periplasmic catalase is plasmid mediated (pO157 plasmid) and provides oxidative protection during infection. *E. coli* O157:H7 fail to show pathogenicity factors seen in other pathogenic *E. coli* and hence cannot be cultured on plates containing bile salts [2]. As a consequence, studies on these hemorrhagic pathogens have been limited and effective means of preventing EHEC have not been elucidated.

*E. coli* O157:H7 was first indicated in food borne illness in the USA in 1982 [1]. In India, the emergence of infections with *E*

*coli* O157:H7 was first reported in 1999 by Pal and colleagues [3]. A comprehensive review of the molecular epidemiology of *E. coli* O157:H7 in India is reported by Wani and colleagues [4]. The recent emergence of this infectious EHEC as a biofilm former has further marked it for increased multiple antibiotic resistances and made it imperative that new strategies should be investigated to curb such infections [1]. One such strategy for novel drug design is targeting metabolic pathways present exclusively in prokaryotes. Vitamin biosynthesis pathways in prokaryotes that are absent in vertebrates provide such an opportune metabolism. The thiamine (Vitamin B1) biosynthesis pathway represents a pathway that has been outlined for *Escherichia coli* as represented in Figure 1 and derived from MetaCyc database [5]. Thiazole (5-methyl-4-(β-hydroxyethyl) thiazole phosphate) and the pyrimidine (4-amino-5-hydroxymethylpyrimidine pyrophosphate) moieties are separately synthesized. The pyrimidine is derived from 5-aminoimidazole ribotide. The thiazole is derived from tyrosine, cysteine, and 1-deoxy-D-xylulose-5-phosphate. These are coupled by the enzyme thiamine pyrophosphorylase kinase (thi E, 2.5.1.3) to form thiamine mono phosphate. A final phosphorylation step results in the formation of thiamine pyrophosphate (THI-PP) which forms the biological active molecule. THI-PP, the metabolically important form of vitamin

B<sub>1</sub> is a cofactor of  $\alpha$ -keto acid dehydrogenase complexes, the glycine cleavage system and enzymes such as transketolase and pyruvate decarboxylase, important in carbohydrate metabolism. Five tightly linked genes have found to be involved in thiamine biosynthesis called as the thi CEF<sub>1</sub>GH cluster [6]. Thi E forms the crucial end point enzyme in the thiamine biosynthesis from different precursors and can be used for as a target for designing drugs against *E. coli* OH: 157 (Figure 1).

In this report, the Thi E enzyme of the TBS pathway was studied in the Enterohemorrhagic *Escherichia coli* OH: 157 using *in silico* approaches. The structural studies on the Thi E have not yet been elucidated. Here, the tertiary structure of the protein is predicted using three different protein structure prediction tools. The best predicted 3D model has been analyzed with respect to its structural features and the putative docking sites of the Thi E enzymes are predicted.

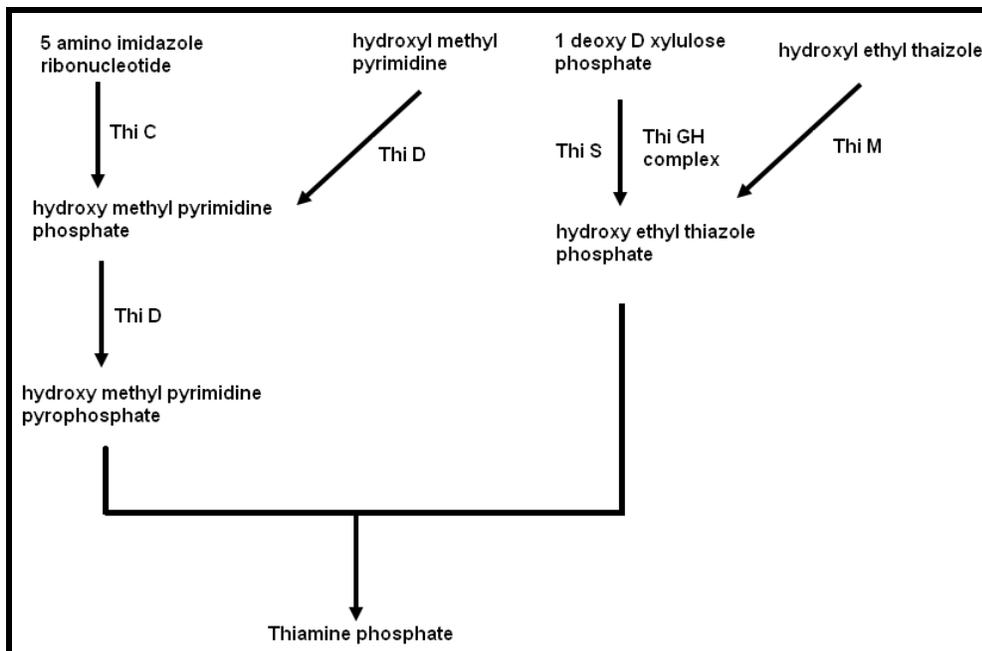
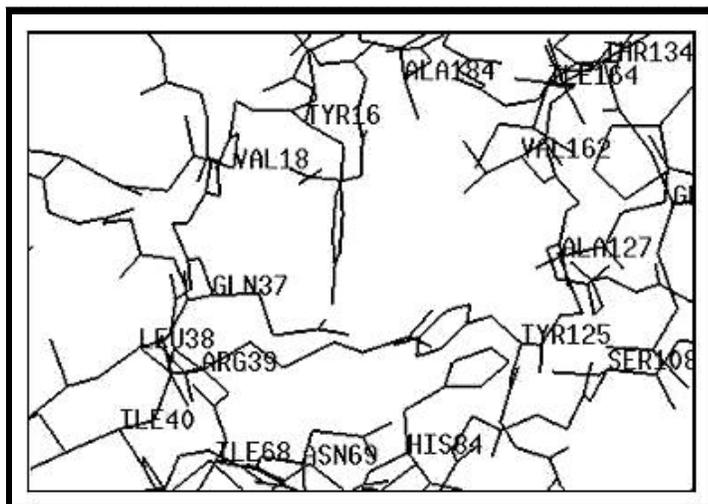


Figure 1: Thiamine biosynthesis pathway for *E. coli*.

Query	12	RLGLYPVVD-----SVQWIERLLDAGVRTLQLRIKDRRDEEVEADVVAALGRRYNARL	66
		+L LY + D V+ + L+ G +Q+RIK+ E+ L R Y+A	
Sbjct	14	KLKLYVITDRRLKPEVESVREALEGGATAIQNRIRKNAPTREMYEIGKTLRQLTREYDALF	73
Query	67	FINDYVURLAIKHQAYGVHVGQEDLQATDLNIRAAGLRLGVSTHDDMEIDVALAARPSYI	126
		F++D +A+ A GV LG ED+ I A L +G S + E A Y+	
Sbjct	74	FVDDRVDVALAVDADGVQLGPEMPEVAKKI-APNLIIGASVYSLEEALAEKKGADYL	132
Query	127	ALGHVFPTQTKQMPQGLEQLARHVERLSDYPTVAIGGISLARAPAVIATGVGSIADV	186
		G VFPT+TK+ + GLE L + VE + P VAIGGI+ A V+ TGV IAV+	
Sbjct	133	GAGSVFPTKTKI-DARVIGLEGLRKRIVESVK-IPVVAIGGINKDNAREVLKTGVVDGIAVI	190
Query	187	SAITQAADURLATAQLLEI	205
		SA+ A D R AT +L +I	
Sbjct	191	SAVMGAEDVRKATELRKI	209

Figure 2: Sequence alignment between target and template.



**Figure 3:** Potential ligand binding pocket residues in Thi E model from *E. coli* OH:157

### Methodology:

#### Sequence retrieval

The FASTA sequence of the target Thi E protein (accession number BAB 38339; length- 211 residues) from *Escherichia coli* O157:H7 was obtained from NCBI Entrez [7].

#### Structure prediction and evaluation

The modeling of the 3 D structure of the protein was performed by three automated homology modeling programs, Geno 3D, Swissmodel and Modeller [8, 9, 10]. The following steps were followed: template structure search using BLASTp [7]. In order to determine homologous sequences to the Thi E enzyme, the FASTA sequence was submitted for NCBI BLASTp analysis. Following BLASTp query, a thiamine phosphate pyrophosphatase from *Pyrococcus furiosus* (PDB ID: 1X13\_A) was selected as template sequence [11]. The template was submitted to Geno3D and Swissmodel automated homology modeling. For Modeller, the template and target sequences were carefully aligned to remove potential alignment errors. The validation for structure models obtained from the three software tools was performed by using PROCHECK [12] and energy minimization performed by Verify3D [13]. The overall stereochemical quality of the protein was assessed by Ramchandran plot analysis [14]. The structures were visualized using Swiss Pdbviewer v 4.0.1.

#### Ligand binding site prediction

An analysis of the ligand binding pockets of the predicted structure was performed using CASTp [15]. As determined by CASTp using a 1.4 Å radius probe, the internal cavity surface volume of the ligand binding sites was calculated.

#### Discussion:

The FASTA sequence of Thi E protein of *E. coli* OH: 157 str Sakai was obtained from NCBI Entrez (Figure 2a). Following BLASTp query, a thiamine phosphate pyrophosphatase from *Pyrococcus furiosus* (PDB ID: 1X13\_A) was selected as template sequence. This sequence showed a highest sequence

homology of 39.3%, with atomic resolution of its X ray crystal structure being 1.7 Å and R value being 0.189. The alignment obtained between Thi E and 1X13\_A is shown in Figure 2. The 1X13\_A structure was used as a template for homology modeling by three different protein prediction tools: Geno3D, Swissmodel and Modeller (9v1). The predicted models were also checked for psi and phi torsion angles using the Ramchandran plots. A comparison of the results obtained from the three different software tools, shows that one of the models generated by Modeller is more acceptable in comparison to that by Geno3D and SwissModel (Table 1 under supplementary material). The molecular visualization program SWISS PDB viewer was used to manipulate the models based on residue interactions, energy minimization and steric hinderance. The best model predicted by Modeller was used for further analysis by PROCHECK [12]. Ramchandran plot analysis shows 91.8% of the residues in the most favored region, 6.6 % in the allowed and 0.5% in the disallowed region. The overall topology of the modeled Thi E from *E. coli* OH: 157 consist of seven strands and nine alpha helices. A TIM beta/alpha barrel fold is observed in the model which is typical of thiamin phosphate synthase superfamily. No homology with any human derived protein was found upon screening the BLASTp analysis. In order to determine the ligand binding pockets in the predicted model, CASTp server was used. A total of 38 ligand binding sites were predicted of which the first pocket was chosen based on the area and the volume of the pocket. The molecular structure of the potential active site predicted by the CASTp server was visualized using Swiss Pdb viewer 4.0.1 (Figure 3).

#### Conclusion:

The study of metabolic pathways for design of suitable targets against infectious agents is an effective strategy. The plethora of enzymes targets in different metabolic pathways remains to be discovered and studied as targets. A molecular model of the phosphoryl thymidine kinase from Enteropathogenic *Escherichia coli* OH: 157 is predicted in this study using three

different homology modeling tools. Thi E forms one of the most conserved parts of the TBS pathway in bacteria while no homology is identified upon blast analysis with humans. The ligand binding sites in the predicted model will provide valuable insights towards inhibitor design.

### Acknowledgment:

Financial assistance from Department of Science and Technology, India is gratefully acknowledged.

### References:

- [01] P. S. Mead *et al.*, *Emerg. Infect. Dis.*, **5**: 607 (1999) [PMID: 10511517]  
 [02] H. Karch *et al.*, *Int J Med Microbiol.*, **6**: 405 (2005) [PMID: 16238016]  
 [03] A. Pal *et al.*, *Indian J. Med. Res.*, **110**: 83 (1999)  
 [04] S. A. Wani *et al.*, *Current Science*, **87**: 1345 (2004)  
 [05] Caspi *et al.*, *Nucleic Acids Research*, **36**: 23 (2008) [PMID: 17965431]  
 [06] G. A. Uhrlich *et al.*, *Appl. Environ. Microbiol.*, **72**: 2564 (2006)  
 [07] <http://www.ncbi.nlm.nih.gov>  
 [08] Combet *et al.*, *Bioinformatics*, **18**: 213 (2002) [PMID: 11836238]  
 [09] K. Arnold *et al.*, *Bioinformatics*, **22**: 195 (2006) [PMID: 16301204]  
 [10] A. Sali and Blundell, *J. Mol. Biol.*, **234**: 779 (1993) [PMID: 8254673]  
 [11] <http://www.rcsb.org/pdb>  
 [12] R. A. Laskowski *et al.*, *J. Biomol. NMR.*, **8**: 477 (1996) [PMID: 9008363]  
 [13] J. U. Bowie *et al.*, *Science*, **253**: 164 (1991) [PMID: 1853201]  
 [14] G. N. Ramachandran *et al.*, *J. Mol. Biol.*, **7**: 95 (1963) [PMID: 13990617]  
 [15] J. Dundas *et al.*, *Nucl. Acids Res.*, **34**: W116 (2006)

Edited by P. Kanguane

Citation: Kaistha & Sinha, *Bioinformatics* 3(6): 240-243 (2009)

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

## Supplementary material

	Modeller	Geno3D	Swissmodel
Model's Energy	-10902.722 KJ/mol	-16039.815 KJ/mol	-9108.529 KJ/mol
Residues in the most FavoredRegion	91.8 %	71.5 %	91.6%
Residues in additionally allowed region	6.6%	23.3%	7.2%
Residues in generously allowed region	1.1%	4.1%	0.6%
Residues in disallowed region	0.5%	1.2%	0.6%

Table 1: Comparative analysis of the validated models from Modeller, Geno3D, and Swiss-model