

Insight from the structural molecular model of cytidylate kinase from *Mycobacterium tuberculosis*

Nitin Kumar Verma^{1,2*} & Balwinder Singh³

¹Department of Bioscience, Shri Ram College, Muzaffarnagar; ²Uttarakhand Technical University, Dehradun; ³Department of Science, Ek Onkar Scholar Degree College, Shahbjnagar, Shahjahanpur; Nitin Kumar Verma - Email: nittinkumarverma@gmail.com; *Corresponding author

Received June 30, 2013; Accepted July 11, 2013; Published July 17, 2013

Abstract:

Mycobacterium tuberculosis is a gram-positive bacterium causes *tuberculosis* in human. H37Rv strain is a pathogenic strain utilized for *tuberculosis* research. The cytidylate mono-phosphate (CMP) kinase of *Mycobacterium tuberculosis* belongs to the family nucleoside mono-phosphate kinase (NMK), this enzyme is required for the bacterial growth. Therefore, it is important to study the structural and functional features of this enzyme in the control of the disease. Hence, we developed the structural molecular model of the CMP kinase protein from *Mycobacterium tuberculosis* by homology modeling using the software MODELLER (9v10). Based on sequence similarity with protein of known structure (template) of *Mycobacterium smegmatis* (PDB ID: 3R20) was chosen from protein databank (PDB) by using BLASTp. The energy of constructed models was minimized and the qualities of the models were evaluated by PROCHECK and VERRIFY-3D. Resulted Ramachandran plot analysis showed that conformations for 100.00% of amino acids residues are within the most favored regions. A possible homologous deep cleft active site was identified in the Model using CASTp program. Amino acid composition and polarity of that protein was observed by CLC-Protein Workbench tool. Exspasy's Prot-param server and CYC_REC tool were used for physiochemical and functional characterization of the protein. Studied of secondary structure of that protein was carried out by computational program, ProFunc. The structure is finally submitted in Protein Model Database. The predicted model permits initial inferences about the unexplored 3D structure of the CMP kinase and may be promote in relational designing of molecules for structure-function studies.

Key words: *Mycobacterium tuberculosis*, Comparative Modeling, MODELLER, Procheck, CMP kinase.

Background:

The genus *Mycobacterium* considers clinically important pathogens such as *M. tuberculosis*, *M. leprae*, *M. bovis* and *M. avium*. Tuberculosis is a disease which spreads from person through the air [1]. Tuberculosis continues a major public health problem of the world; it has been further aggravated by the advent of HIV-AIDS. *M. tuberculosis* (Mtb) is a facultative intracellular parasite which grows at slow rate. In different environmental condition Mtb exposed rigorous stages of disease during infection. Mtb replicates inside the macrophages, where the condition is mostly hostile for most bacteria. It can also multiply extracellularly in the open lung cavities which are found during the late stages of the disease [2]. In the last decades, the cost of research and development in the pharmaceutical industry has been rising steeply and steadily,

but the significant time is required for bringing a new product to market remains around ten to fifteen years [3]. Tuberculosis has always been a major killer, making in India. One of the natural testing ground for the highly promising *M. bovis*-BCG, the only vaccine available for TB even today. The near-completed failure of the BCG vaccination programmed in India [4]. The two main drug resistant are MDR (Multidrug-resistant) and XDR (extensively drug-resistant), first line anti-TB drug are rifampicin and isoniazid in case of Multidrug-resistant tuberculosis (MDR-TB)[6]. MDR-TB has additional resistance to fluoroquinolone and second line injectable antibiotics are amikacin, kanamycin or capreomycin, in case of XDR [5, 6]. More recently, a survey of XDR-TB cases during 2000-2004, in this period 17,690 TB isolate, 20% were MDR and 2% were XDR.

Mycobacterium tuberculosis H37Rv was first isolated in 1905, has remained pathogenic and this strain is most widely used in *tuberculosis* research. The complete annotated genome sequence of H37Rv strain was published in 1998 [7]. CMP kinase is the key enzyme in the nucleotide metabolism, which is connected to the family of nucleoside monophosphate kinase (NMP) [8]. Substrate specificity studied on recombinant human UMP/CMP kinase (primidine nucleoside monophosphate kinase) has shown that UMP and CMP are far better substrates than dCMP [9]. Bacterial cytidylate kinase or cytidine monophosphate kinase (CMP kinase) catalyses the phosphoryl transfer from ATP to CMP and dCMP, resulting in the formation of nucleoside diphosphates. CMP kinase has been shown to be necessary for bacterial growth in *Streptococcus pneumoniae* [10]. The *Staphylococcus aureus* CMP kinase crystal structure has been recently resolved [11]. In the protein structure of *Escherichia coli* CMP kinase resembles those of other NMP kinases sharing common features such as a central five-stranded β -sheet connected by α -helices, a fingerprint sequence of Glu-x-x-Gly-x-Gly-Lys (P-loop), and an anion hole in the central cavity for substrate binding [12]. However, *E. coli* CMP kinase contains an insertion of 40 residues and a short LID domain, differing from other NMP kinases [12]. These differences could allow the rational design of drugs specific for CMP kinase of pathogenic bacteria, such as *M. tuberculosis*. However, there is no information available, which describes the structural properties of CMP kinase from *M. tuberculosis*. In this work, we re-construct the structure of CMPK from Mtb H37Rv in order to elucidate the structural properties. We acquire better result in comparison to previous study.

Methodology:

Sequence Retrieval and Template selection

The protein sequence of cytidylate kinase enzyme of "*Mycobacterium tuberculosis*" was retrieved from NCBI database (National Center for Biotechnology Information) which have 230 amino acid residues, accession number is NP_216228 for modeling. Comparative modeling usually starts by searching the PDB (Protein Data Bank) of known protein structure using the target sequence as a query. This search is generally done by comparing the target sequence with the sequence of each of the structure in the database. The protein sequence of *Mbt*CMPK worked as a target sequence. For identification of suitable template BLASTp was performed against PDB, select templates for generating the putative 3D structures.

Comparative Modeling

The method of comparative modeling requires identification of homologous sequences with known structure. The first step of comparative modeling scanning and selection of template protein structure using the target sequence as the query [13]. BLASTp generally used for template selection procedure [14]. The BLAST results outputted potential template for modeling is 3R20, the high degree of primary sequence identity (71.40%) between *Mt*CMPK (target) and *Mycobacterium smegmatis* indicates that this crystallographic structure was good model to be used as template for *Mt*CMPK enzyme. The alignment of the *Mt*CMPK target and *Mycobacterium smegmatis* CMPK is shown in (Figure 1). A total of 5 models were generated using program MODELLER 9v10 [15]. MODELLER implements comparative protein modeling by satisfaction of spatial restraints.

Analysis of the model

Modeller generated the several models for the same target, the best model can be selected for further analysis. We evaluated the model with the lowest value of the Modeller objective function and PROCHECK statistics. The overall stereochemical quality of the models for *Mt*CMPK enzyme using Ramachandran's plot calculation computed with PROCHECK [16] program, available on NIH (National Institute of Health) server (http://nihserver.mbi.ucla.edu/SAVES_3/Procheck/) and the final structure was subsequently checked by VERIFY-3D graph available at NIH server (http://nihserver.mbi.ucla.edu/SAVES_3/Verify_3D/). The summarized data of all five models is in Table 1 (see supplementary material).

MTB_CMP	1	--MSRLSAA-VVAIDGPAGTGKSSVSRRLARELGARFLDTGAMNRIIVTLA	47
3R20_A	1	GPGSMVSGSLVVAIDGPAGTGKSSVSRGLARALGARYLDTGAMNRIATLA	50
MTB_CMP	48	VLKAGADPSDIAAVETIASTVQMSLGYDPDGDSCYLAGEVDSVEIRGDAV	97
3R20_A	51	VLKAGADLTPAAIEKAAADAIEVGSDDPVDAAFLAGEVDSSEIRGDAV	100
MTB_CMP	98	TRAVSAVSSVPAVTRRLVELQRTMAEGPGSIVVEGRDITGVVFPDAPVKI	147
3R20_A	101	TGAVSAVSAVPAVTRRLVDIQRLKATEGGRVVEGRDITGVVLPDADVKI	150
MTB_CMP	148	FLTASAETRARRRMAQNVAAAGLADDYDGLADVRRRHLSDTRAVSPLQA	197
3R20_A	151	FLTASAEEERARRRMAQNVANGLPDDYATVADVQRRHLSDTRPVSPLRA	200
MTB_CMP	198	AGDAVIVDTSMDTEAEVVAHLLLELVTRRSEAVR	230
3R20_A	201	ADDALVVDTSMDQAQVIAHLLLDLVTAQAQARR	233

Figure 1: Sequence Alignment of *Mycobacterium tuberculosis* (NP_216228) with *Mycobacterium smegmatis* (PDB ID: 3R20). The alignment was performed using EMBOSS.

Visualization and structural analysis

Visualization and Structural analysis of final protein model were carried out with Visual Molecular Dynamics (VMD) [17] software. Structural alignment and RMSD calculation with several models and template were carried out with VMD.

Active site analysis

After the final model was build, the possible deep cleft active sites of *Mt*CMPK were explored applying CASTp program [18]. (<http://sts.bioengr.uic.edu/castp>)

Physicochemical characterization

For physicochemical characterization, theoretical pI (isoelectric point), molecular weight, -R and +R (total number of positive and negative residue), EI (extinction coefficient), II (instability index), AI (aliphatic index), and GRAVY (grand average hydropathy) were computed using the ExPASy's ProtParam server for set of proteins (<http://us.expasy.org/tools/protparam.html>). The results are showing in Table 2 (see supplementary material).

Functional characterization

CYC_REC (<http://sunl.softberry.com/berry.phtml?topic>) was used to locate "SS bond" between the pair of cystein residue, if present. The tool yields position of cystein residue, total number of cysteins present and pattern, if present of pairs in the

protein sequence as output **Table 3** (see **supplementary material**).

Secondary structure prediction

ProFunc was employed for calculating the secondary structure features of cytidylate kinase protein sequence [19].

Submission of the modeled protein in protein model database (PMDB)

The model generated of *Mycobacterium tuberculosis* cytidylate kinase protein was successfully submitted in Protein model database without any stereochemical errors. The submitted model can be accessed via their **PMID: PM00779121**.

Results & Discussion:

Model building, refinement and evaluation

In earlier work, molecular modeling and dynamic studies of *MtCMPK* H37Rv have been done. In this study, the primary sequence identity between *MtCMPK* (target) and *E.coli* CMPK indicates that ~40% sequence identity (**Figure 2**). Medium accuracy model, obtained with a template-target sequence identity of 30-50% [20]. The sequence analysis result based on BLASTp, there were 220 amino acids only 88 amino acids had been matched. A total 1000 models had been generated for each binary complex and the final model selected based on MODELLER objective function [21]

MTB_CMP	1	MSRLSAAVVAIDGPAAGTGKSSVSRRLARELGARFLDTGAMYRIVTLAVLR	50
ECOLI_CMP	1	MTAI-APVITIDGPGSGAGKGLCKAMAEALOMHLLDSGAIYRVLALAAH	49
MTB_CMP	51	AGADPSDIAAVETIASTVQMSLGYDPDGDSCYLAGEDVSVVEIRGDAVTRA	100
ECOLI_CMP	50	HHVDVASEDALVPLASHLDVRFVSTNGNLEVLLEGEDVSGEIRTQEVANA	99
MTB_CMP	101	VSAIVSVPVAVRTRLVELQRTMAEGPGSIVVEGRDITGVVFPDAPVKIFLT	150
ECOLI_CMP	100	ASQVAAFPRVREALLRRRAFRELPG-LIADGROMGTVFPDAPVKIFLD	148
MTB_CMP	151	ASAETRARRRINAQNAAGLADDYDGLADVRRDRHLDSTRAVSLQAAGD	200
ECOLI_CMP	149	ASSEERARRMLQLQEKGFVNFERLLAEIKERDORRWRVAVPLVPAAD	198
MTB_CMP	201	AVIVDTSDMTEAEVVAHLLVTRSEAVR	230
ECOLI_CMP	199	ALVLDSTLSEIQVIEKALQYAROKLALA-	227

Figure 2: Sequence Alignment for *EcCMPK* and *MtCMPK*. The sequence identity between *EcCMPK* and *MtCMPK* is ~40%.

In the current research work, experimental structure of cytidylate kinase of *Mycobacterium tuberculosis* is not yet available in PDB (Protein Data Bank), comparative modeling approach was used in order to derive their structure. BLAST scanning results had shown higher similarity with crystallographic structure of 3R20, while the template was selected on the basis of higher sequence identity. It shows 71.40% sequence identity, with 167 conserved residue and 83.30% sequence similarity. Comparative modeling predicts the 3-D structure of cytidylate kinase model of given protein sequence (target), based primarily on this alignment to the template. The model was also checked for phi and psi torsion angles using the Ramachandran plots. Altogether more than 90% of the residues were found to be in favored and allowed

regions, which validate the quality of homology models. The modeled structures were also validated by other structure verification server such as Verify 3D. Verify 3D assigned a 3D-1D score of >0.2 and residue result is 96.10%. Visualization and analysis of the model using Visual Molecular Dynamics (VMD) reveals that there are no steric hindrances between the residue and thus modeled structure are stable. Structure-structure superimposition was done in order to calculate Root Mean Square Deviation (RMSD) between the target and template sequence. RMSD value is 0.1929 nm. This implies good quality of the modeled structure. We represent modeled structure of cytidylate kinase (**Figure 3**).

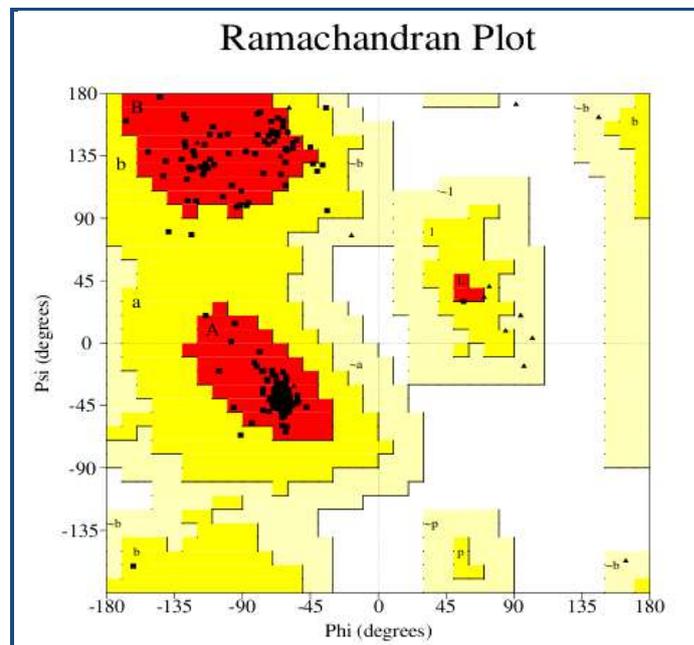


Figure 3: Ramachandran Plot of the cytidylate kinase (*Mycobacterium tuberculosis*) by PROCHECK

Physiochemical characterization

The physiochemical properties videlicet, theoretical isoelectric point (Ip), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index, grand average hydropathy (GRAVY) were calculated using the Expasy's ProtParam tool in **Table 2** (see **supplementary material**). The calculated pI value for cytidylate kinase indicated their acidic nature. The computed pI will be helpful for developing buffer system for purification by isoelectric focusing method. An extinction coefficient value for the cytidylate kinase protein at 280 nm is 5960 M⁻¹cm⁻¹. On the basis of instability index, protein is unstable due to instability index is greater than 40. The aliphatic index (AI) which is defined as the relative volume of a protein occupied by aliphatic side chain is regarded as the positive factor for the increase of thermal stability of the protein. The very high aliphatic index of cytidylate kinase protein which indicates that protein may be stable for a wide range of temperature. The GRAVY index of protein which indicate no interaction with water because protein has positive score.

Functional characterization

The results of primary analysis suggest that the *Mycobacterium tuberculosis* cytidylate kinase under study was hydrophobic in

nature. As percentage of Cysteine (C) is very low in cytidylate kinase under study, it have not disulphide bond linkages indicated by CYS_REC (Cys_81, score = -2.2). The extensive hydrogen bonding may provide stability to this protein in absence of disulphide bonds. Cytidylate kinase have high percentage of alanine (A), leucine (L), glutamic acid (E). As these amino acids have high-helical forming propensities, alpha helix and beeta strands are dominant in that protein. Ten alpha-helices surround the beta-ribbon structure. The helical region are composed of the residue 19-30, 38-51, 58-66, 89-91, 95-105, 109-120, 134-138, 153-167, 173-189 and 211-225. The seven beta-ribbon are composed residues 7-12, 33-37, 69-75, 78-84, 127-132, 145-150 and 202-206. In (Figure 4), it shows a schematic drawing of the MtCMPK consist of 230 amino acid with a predicted molecular mass of 24,177.29 Da and a theoretical pI value of 5.05.

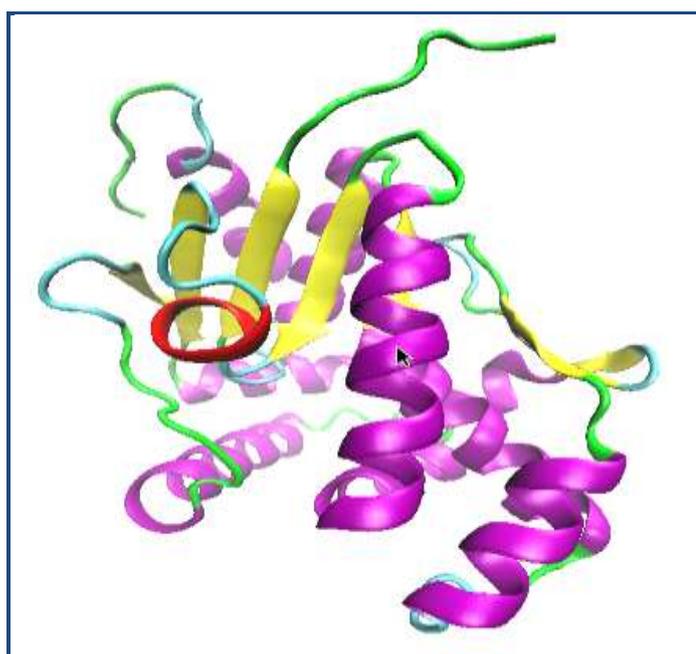


Figure 4: Predicted protein structure of cytidylate kinase (*Mycobacterium tuberculosis*)

Submission of modeled proteins in PMDB

The modeled structure of cytidylate kinase from *Mycobacterium tuberculosis* successfully deposit in Protein Model Database (PMDb). The submitted structure of protein is **PMDB ID: PM0079121**.

Conclusion:

In this study, we have constructed a 3-D model of cytidylate kinase from *Mycobacterium tuberculosis*, using the comparative modeling approach. Different parameters such as isoelectric point, molecular weight, total number of positive and negative residue, extinction coefficient, instability index, aliphatic index

and grand average hydropathy (GRAVY) were computed for this protein in order to determine their physicochemical characteristics. In cytidylate kinase was found very low percentage in amino acid cysteine, and therefore lack presence of Disulphide Bridge as also inferred from analysis of cys_res result. In the absent of disulphide bond, extensive hydrogen bonding is believed to be responsible for stability of the protein. Polarity studies using CLC protein work bench tool confirmed cytidylate kinase to be hydrophobic in nature. Secondary structure analysis predicts alpha-helices are dominant in the protein. The modeled structure can be accessed through protein modeled database PMDB via there PMID. This structure will serve as a cornerstone for functional analysis through experimentally derived crystal structure.

Acknowledgment:

The authors are grateful to the Department of Bioscience, Shri Ram College, Muzaffarnagar for providing the necessary facilities. Authors are grateful to Bioinformatics editorial board members and Bioinformatics team of reviewers who have helped to bring quality to this manuscript. We also thank to Mr. Ravi datta sharma and Vibha for their generously support and encouragement.

References:

- [1] Asif SM *et al. Bioinformatics*. 2009 **4**: 245 [PMID: 20975918]
- [2] Dannenberg AM Jr *et al. Ameri Soc Microbiol*. 1994 **459**
- [3] <http://www.roche.com/>
- [4] Bloom BR *et al. Ameri Soci Microbiol*. 1994 **531**
- [5] <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5543a4.htm>
- [6] WHO XDR-TB, *Wkly Epidemiol Rec*. 2006 **81**: 430 [PMID: 17096498]
- [7] Cole ST *et al. Nature*. 1998 **393**: 537 [PMID: 9634230]
- [8] Leipe DD *et al. J Mol Biol*. 2003 **333**: 781 [PMID: 14568537]
- [9] Liou JY *et al. Cancer Res*. 2002 **62**: 1624 [PMID: 11912132]
- [10] Yu L *et al. Protein Sci*. 2003 **12**: 2613 [PMID: 14573872]
- [11] Dhaliwal B *et al. Acta Crystallogr Sect F Struct Biol Cryst Commun*. **62**: 710 [PMID: 16880539]
- [12] Briozzo P *et al. Structure*. 1998 **6**: 1517 [PMID: 9862805]
- [13] Westbrook J *et al. Nucleic Acids Res*. 2002 **30**: 245 [PMID: 11752306]
- [14] <http://blast.ncbi.nlm.nih.gov/Blast>
- [15] Marti-Renom MA *et al. Annu Rev Biophys Biomol Struct*. 2000 **29**: 291 [PMID: 10940251]
- [16] Laskowski RA *et al. J Appl Cryst*. 1993 **26**: 283
- [17] Humphrey W *et al. J Mol Graph*. 1996 **14**: 33 [PMID:8744570]
- [18] Dundas J *et al. Nucleic Acids Res*. 2006 **34**: W116 [PMID: 16844972]
- [19] <http://www.ebi.ac.uk/thornton-srv/database/pdbsum>
- [20] Kherraz *et al. Bioinformatics*. 2011 **6**: 115 [PMID: 21584187]
- [21] Caceres RA *et al. J Mol Model*. 2008 **14**: 427 [PMID: 18343960]

Edited by P Kanguane

Citation: Verma & Singh, *Bioinformatics* 9(13): 680-684 (2013)

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:

Table 1: Summary of PROCHECK result for predicted protein constructed by MODELLER

Structure Name	Core	Additional Allowed region	Generously Allowed region	Disallowed region
CMP.B99990001	91.6	8.4	0.0	0.0
CMP.B99990002	95.1	4.9	0.0	0.0
CMP.B99990003	92.6	7.4	0.0	0.0
CMP.B99990004	94.6	5.4	0.0	0.0
CMP.B99990005	94.1	5.9	0.0	0.0

Table 2: Physiochemical characters as predicted by Expasy's prot-param program

Seq. Len.	M.Wt.	pI	-R	+R	Extinction Coefficient	Instability Index	Aliphatic Index	GRAVY
230	24177.20	5.05	32	26	5960	44.35	98.83	0.055

Table 3: Presence of disulphide (ss) bonds as predicted by Cys_Rec

Name of Protein	CYS_REC	Score
Cytidylate kinase	Cys_81	-2.2