Journal of Advanced Bioinformatics Applications and Research
ISSN 0976-2604.Online ISSN 2278 – 6007
Vol 4, Issue 1, 2013, pp11-14
http://www.bipublication.com

COMPARATIVE MODELLING OF EMBB PROTEIN OF *Mycobacterium tuberculosis* AND STRUCTURE REFINEMENT

Lalit R. Samant[1], Sudhir Kumar Pal[2] and Upendra Bhele[2]

[1]srujan Biotech And Renewable Energy Solutions, Dombivali, Thane, MH, India
[2]Dr. DY Patil University, Departement of Biotechnology and Bioinformatics,
Plot no.-50, Sector-15, CBD Belapur, Navi Mumbai-400614
Corresponding author : samantlalit@gmail.com, M- +91 91 72 300 601

[Received-01/09/2012, Accepted-21/02/2013]

ABSTRACT:

Mycobacterium tuberculosis is causative agent of the one of the deadliest disease, the bacteria showing resistance for the most of the antibiotics and hence the generation of new drug is needed, we have focused on the embb gene protein for the modelling which plays a vital role in manifestation of the disease, the structure was modelled and deposited to the PMDB with PMDB ID: PM0078399

INTRODUCTION

Mycobacterium tuberculosis is an extremely successful pathogen that demonstrates the capacity to modulate its host both at the cellular and tissue levels. At the cellular level, the bacterium enters its host macrophage and arrests phagosome maturation, thus avoiding many of the microbicidal responses associated with this phagocyte. *Mycobacterium tuberculosis* (Mtb) remains one of the most pernicious of the infectious diseases borne by mankind (1) Most cell biological studies of intracellular pathogens proceed by either immunofluorescent co-localization of known 'marker' proteins, or through the use of transfection of host cells with proteins tagged with reporters such as green fluorescent proteins (GFPs). These approaches have been and continue to be extremely informative (2-10).

MATERIALS AND METHODS[23] :-
The sequence was taken from Swissprot for Homology Modeling whose structure was unavailable on the RCBS PDB.

>tr|Q745M3|Q745M3_MYCPA EmbB OS=Mycobacterium paratuberculosis GN=embB PE=4 SV=1
MSVSTVGGDVRVTRWVATIAGLIGFVLSVATPLLPV
VQTTATLNWPQGGQLNSVTAPLIS
LTPVDLTATVPCSSVRDLPPEGGVILSTGPKKGKDAA
LNALFVVAHGKRVDVTDRNVVIA
SASRDQVAGAGCSRIEITSTRAGTFATFVGLTDPAGK
PLGGGFPDPNLRPQIVGVFTDLT
GPSPAGLKLSATIDTRFSTTPTTLKLAAMVTAILATIV
ALVALWRLDQLDGHRMRRLIPA
NWRTFTLADVAVIFGFVLWHVIGANSSDDGYILGMA
RVADRAGYMSNYFRWFGSPEDPFG
WYYNLLALMTHVSDASLWMRLPDLFAGIVCWLLLS
REVLPRLGPAVAASRPANRAAGMVL
LTAWMPFDNGLRPEPIIALGSLVTYVLIERSMRYSRL
TPAALAVITAAFTLGVQPTGLIA
VAALVAGGRPILRILVRRHRVVGTWPLVAPMLAAG
TVILTVVFADQTLATVLEATRIRTA
IGPSQAWYTENLRYYYLILPTVDGSLSRRFGFLITAL
CLFTAVFIMLRRKRIPGVARGPA
WRLMGVIFGTMFFLMFTPTKWVHHFGLFAAVGAA
MAALTTVLVSPAVLRWSRNRMAFLAA
LLFMMALCFATTNGWWYVSSYGVPFNSTMPKIGGI
TVSTVFFAMFVAAALYAIGLHFASR
EHGDGRLARALTAAPVPLAAGFMALVFIASMVAGIV
RQYPTYSNAWDNLREFSGGCGLAD
DVLVEPDSNVGYMTPLGGDYGPLGPLGGQHPVGFS
PNGVPEHTVAEAIRITPNQPGTDYD

WDAPTKLSAPGINGSTVPLPYGLDAARVPLAGSYTT
GAQQQSRLTSAWYRLPAPDDGHPL
VVVTAAGKIAGNSVLHHHTDGQTVVLEYGRPGPGG
DIVPAGRLVPYDLYGEQPKAWRNLR
FARSDMPADTVAVRVVAEDLSLTPEDWIAVTPPRVP
EMRSLQEYVGSTQPVLMDWAVGLA
FPCQQPMLHVNGVTEIPKFRITPDYTAKKMDTDTWE
DGTNGGLLGITDLLLRAHVMSTYL
SHDWGRDWGSLRRFETIADAHPAQLDLGTATRTGW
WSPGPIRIKP

2.1 Physiochemical characterization[15-18]:
Physiochemical characterization, theoretical pI (isoelectric point),molecular weight, -R and +R (total number of positive and negative residues),EI (extinction coefficient) , II (instability index , AI (aliphatic index) and GRAVY (grand average hydropathy) were computed using the Expasy'sProtParam server for MS . protein(http://us.expasy.org/tools/protparam.html). The results are shown in Table 1.

Table 1: Physiochemical characters as predicted by Expasy's prot-param program

Length	M.wt	pI	-R	+R	Extinction Coefficient	Instability Index	Aliphatic Index	GRAVY
1065	114626.9	9.23	38	34	196095	37.69	97.83	0.276

2.2 Functional characterization:
Cysteine Recognition server yielded the position of cysteins, total number of cysteins present and pattern, MS protein under study showed absence of disulphide bonds. Secondary structure prediction: [11, 12] was employed for calculating the secondary structural features of protein Sequence. . Trans membrane region predation was carried out using HMMTOP server Table The results are presented in Table 4 and 5[13,14].

2.3 **Model building and evaluation:**
The three dimensional structures of proteins were modeled using Phyre 2 and Swiss workbench. [19]. Quality of generated models was evaluated with by QMEAN 4 ANNOLEA, GROMOS analysis [20-26].

2.4 **Model Refinement**:
The modelled structures were having some anomalies in the side chain and folds which have been corrected by software FOLDX. The refined structure can be seen in the figure.

3. RESULT AND DISCUSSION:

Physiochemical characterization :
From Table 1. We can conclude ; The physiochemical parameters viz., theoretical isoelectric point (Ip), molecular weight, total number of positive and negative residues, extinction coefficient, half-life, instability index, aliphatic index and grand average hydropathy (GRAVY) were computed using the Expasy's ProtParam tool. The computed isoelectric point will be useful to for developing buffer system for purification by isoelectric focusing method. Extinction coefficient values for protein at 280 nm is 196095 M-1cm-1 indicating the presence of higher concentration of Tyr and Trp and assuming all pairs of Cys residues form cystines. The instability index (II) is computed to be 37.69 This classifies the protein as stable.. The very low GRAVY index of protein infers that these proteins could result in a better interaction with water.

Functional characterization:

The CYS_rec results shows the presence of two disulphide bonds indicates the stability of protein might be increased due to this along with non covalent interactions. The a.a Compostion shows higher value Lys, Leu serine shows the amino acids have high helix-forming propensities, alpha helix are dominant in these proteins (Table 3) . Secondary structure prediction: [11,12] was employed for calculating the secondary structural features of protein Sequence. . Trans membrane region predation was carried out using HMMTOP server Table The results are presented in Table 4 and 5[13-15,7].

Table 2: Cys Rec Result

6 cysteines are found in positions	37 41 43 53 295 328
The most probable pattern of pairs	37-43, 53-295

Table 3.: Amino Acid Compostion

Amino acid composition:

Ala (A)	12	3.6%
Arg (R)	9	2.7%
Asn (N)	30	8.9%
Asp (D)	17	5.0%
Cys (C)	6	1.8%

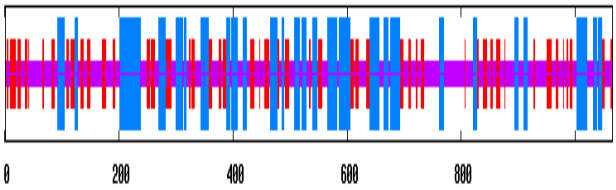
Gln (Q)	10	3.0%
Glu (E)	21	6.2%
Gly (G)	15	4.4%
His (H)	9	2.7%
Ile (I)	19	5.6%
Leu (L)	55	16.3%
Lys (K)	25	7.4%
Met (M)	4	1.2%
Phe (F)	14	4.1%
Pro (P)	18	5.3%
Ser (S)	31	9.2%
Thr (T)	11	3.3%
Trp (W)	1	0.3%
Tyr (Y)	17	5.0%
Val (V)	14	4.1%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

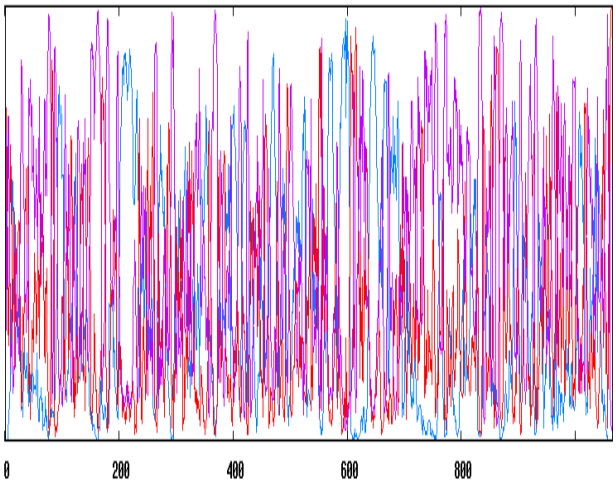
Table 4: Trans membrane Region prediction By HMMTOP(13)

Protein	MS Protein
Length	1065
N-terminus	IN
Number of transmembrane helices:	13
Transmembrane helices: NO	15-37 207-224 245-262 315-334 355-371 402-426 447-464 510-527 538-556 565-582 595-612 633-655 676-695

Table 5 : Secondary structure prediction by GOR IV and SOPMA (11,12)

Alpha helix (Hh) :	147	28.92%	37%
3_{10} helix (Gg)	0	0.00%	0.00%
Pi helix (Ii)	0	0.00%	0.00%
Beta bridge (Bb)	0	0.00%	0.00%
Extended strand (Ee) :	26	21.31%	13.90%
Beta turn (Tt) :	0	0.00%	0.00%
Bend region (Ss) :	0	0.00%	0.00%
Random coil (Cc) :	165	49.77%	45.73%
Bend region (Ss) :	0	0.00%	0.00%
Ambigous states :	0	0.00%	0.00%
Other states :	0	0.00%	0.00%



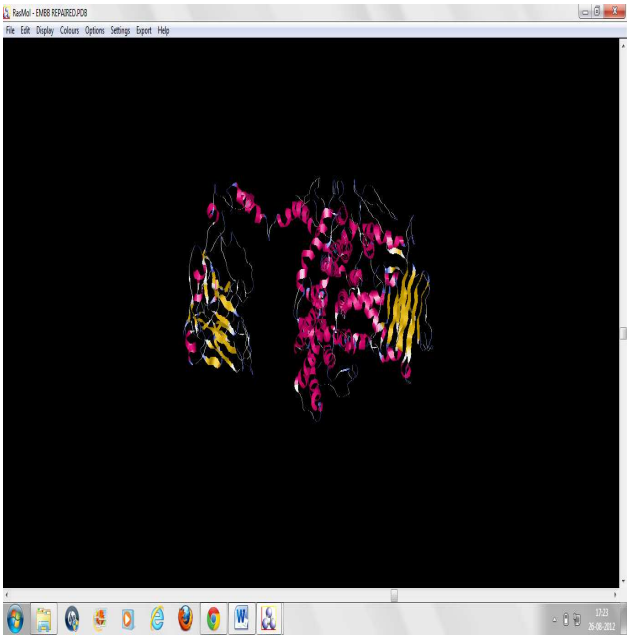


Modelling and Structure validation By QMEAN (19-26):-

A 3D model of protein has been built using the 3D structure 3PTY chain 'A' as template. This template shares 17.1% identities with your query sequence (using the ALIGN program) USING phyre 2 server. The refined structure was validated using QMEAN and the structure was good from the predictions.

Model name: embB_repaired.pdb	embB_repaired.pdb
C_beta interaction energy:	326.72
All-atom pairwise energy:	4087.08
Solvation energy:	48.57
Torsion angle energy:	71.28
Secondary structure agreement:	76.9%
Solvent accessibility agreement:	66.2%
Total QMEAN-score: (estimated model reliability between 0-1)	0.381

Figure:- Refined Structure of EMBB using FOLDX software



CONCLUSION:-

The modelled refined structure has been deposited to PMDB database with PMID: PM0078399 which can be further used for study in future the work is under progress which might result in a new drug candidate for fighting against *M. Tuberculosis*

REFRENCES:-

1. Russell DG, Barry CE 3rd, Flynn JL. Tuberculosis: what we don't know can, and does, hurt us. Science. 2010; 328:852–856. [PubMed: 20466922]
2. Cardoso CM, Jordao L, Vieira OV. Rab10 regulates phagosome maturation and its overexpression rescues Mycobacterium-containing phagosomes maturation. Traffic. 2010; 11:221–235. [PubMed: 20028485]
3. Philips JA, Porto MC, Wang H, Rubin EJ, Perrimon N. ESCRT factors restrict mycobacterial growth. Proc Natl Acad Sci USA. 2008; 105:3070–3075. [PubMed: 18287038]
4. Roberts EA, Chua J, Kyei GB, Deretic V. Higher order Rab programming in phagolysosome biogenesis. J Cell Biol. 2006; 174:923–929. [PubMed: 16982798]
5. Roberts EA, Deretic V. The Mycobacterium tuberculosis phagosome. Methods Mol Biol. 2008; 445:439–449. [PubMed: 18425467]
6. Seto S, Matsumoto S, Ohta I, Tsujimura K, Koide Y. Dissection of Rab7 localization on Mycobacterium tuberculosis phagosome. Biochem Biophys Res Commun. 2009; 387:272–277. [PubMed: 19580780]
7. Seto S, Matsumoto S, Tsujimura K, Koide Y. Differential recruitment of CD63 and Rab7-interacting-lysosomal-protein to phagosomes containing Mycobacterium tuberculosis in macrophages. Microbiol Immunol. 2010; 54:170–174. [PubMed: 20236428]
8. Sweet L, Singh PP, Azad AK, Rajaram MV, Schlesinger LS, Schorey JS. Mannose receptordependent
9. delay in phagosome maturation by Mycobacterium avium glycopeptidolipids. Infect Immun. 2010; 78:518–526. [PubMed: 19841083]
10. Vergne I, Chua J, Singh SB, Deretic V. Cell biology of mycobacterium tuberculosis phagosome. Annu Rev Cell Dev Biol. 2004; 20:367–394. [PubMed: 15473845]
11. GOR secondary structure prediction method version IV, J. Garnier, J.-F. Gibrat, B. Robson, Methods in Enzymology,R.F. Doolittle Ed., vol 266, 540-553, (1996)
12. Geourjon, C. & Deléage, G., SOPMA: Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments., Cabios (1995) 11, 681-684
13. G E Tusnády and I Simon J. Mol. Biol., **1998**, 283, 489-506.
14. G E Tusnády and I Simon Bioinformatics, **2001**,17, 849-850.
15. NORS: Liu J, Rost B NORSp: Nucleic Acids Research **2003**, 31, 3833-3835.
16. Ceroni, A. Passerini, A. Vullo and P. Frasconi, Nucleic Acids Research, **2006**, 34, 177--181
17. http://linux1.softberry.com
18. http://rostlab.org/papers/1999_globe/paper.html
19. A Bairoch, P Bucher ∧ K Hofmann, Nucleic Acids Research, **1997**, 25, 217-221.
20. Phyre 2 server: http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index
21. P Benkert, M Biasini, T Schwede, Bioinformatics, **2011**,27, 343-350.
22. T Castrignanò, De Meo PD, D Cozzetto, IG Talamo, A Tramontano, Nucleic Acids Research, **2006**, 34, 306-309.
23. Vidhya VG, Anusha Bhaskar, Parthiban Purushothaman, J. Comput. Method. Mol. Design, 2011, 1, 22-27.
24. Benkert, P, Tosatto, S.C.E. and Schomburg, D. 2008. "QMEAN: A comprehensive scoring function for model quality assessment." Proteins: Structure, Function, and Bioinformatics, 71(1):261-277.
25. Benkert P, Kuenzli M, Schwede T. (2009). "QMEAN Server for Protein Model Quality Estimation."Nucleic Acids Res. 2009 Jul 1;37(Web Server issue):W510-4.
26. Benkert P, Biasini M, Schwede T. (2011). "Toward the estimation of the absolute quality of individual protein structure models." Bioinformatics, 27(3):343-50.