In silico Comparative Modeling of PapA1 and PapA2 Proteins Involved in *Mycobacterium Tuberculosis* Sulfolipid-1 Biosynthesis Pathway

Rana Adnan Tahir[1*], Sheikh Arslan Sehgal[1], Ambreen Ijaz[2]

[1]*Department of Bioinformatics and Biotechnology*
International Islamic University Islamabad
H-10 Sector Islamabad, Pakistan
E-mails: ranaadnantahir@yahoo.com, arslansehgal@yahoo.com

[2]*Department of Bioinformatics & Biotechnology*
Government College University Faisalabad
Allama Iqbal Road, 38000, Pakistan
E-mail: ambreen_7868@yahoo.com

[*]*Corresponding author*

Received: December 31, 2011 **Accepted: September 12, 2012**

Published: September 15, 2012

Abstract: *Tuberculosis is one of the most serious health problems, as globally; around 2 billion or one third of the world's total population has been infected with Mycobacterium tuberculosis. Mycobacterium tuberculosis is a unique among bacterial pathogens in that it displays a wide array of complex lipids and lipoglycans on its cell surface. One such glycolipid, sulfolipid-1 (SL-1), is the most sulfatide, consists of a trehalose core, four fatty acyl groups, and a sulfate ester. Several proteins involved in SL-1 biosynthesis have been identified, the enzymes that acylate the T2S core to form SL1278 and SL-1, and the biosynthetic order of these acylation reactions, are unknown. Here we studied the in silico identification of PapA2 and PapA1, proteins responsible for the sequential acylation of T2S to form SL1278 and are essential for SL-1 biosynthesis, by applying different bioinformatics tools. Benchmark, of 3 different homology modeling programs Modeller, Swiss-Model (Deep View), and ESyPred3D, has been performed used to transform the alignment to a 3D model. The 3D structures of targeted proteins were evaluated by evaluation tools, ANOLEA and Verify3D. It is concluded that in SL-1 biosynthesis pathway, PapA1 and PapA2 proteins could be used as drug target, drug lead design and to find out the other proteins involved in this pathway that not yet have been identified and may be used to the cure of tuberculosis infection.*

Keywords: *PapA1 and PapA2, Mycobacterium Tuberculosis, Comparative modeling, Bioinformatics analysis.*

Introduction

Tuberculosis is a major global cause of death and infectious disease and around 2 billion people or one third of the world's total population, are believed to be infected with tubercle bacilli (TB) [37].Today the principle cause of human tuberculosis is *Mycobacterium tuberculosis* [14]. *Mycobacterium tuberculosis* is a gram-positive bacterium with the G+C rich genome. The thick *Mycobacterium tuberculosis* cell wall consists of numerous glycolipids that are essential for many of the characteristics that distinguish Mycobacterial pathogenesis [5, 28].

Genes involved in the synthesis and export of surface-exposed lipid virulence factors such as Phthiocerol Dimycocerosate (PDIM) and Sulfolipid-1 (SL-1) are required for bacterial growth and virulence in mice [8, 9]. Surface-exposed lipids provide protection against host induced

damage [26, 27]. A family of cell surface sulfated lipids was identified in *Mycobacterium tuberculosis* extracts and correlated to strain virulence [10, 23]. Sulfolipid-1 (SL-1) is the most abundant sulfatide, consists of a trehalose core, four fatty acyl groups, and a sulfate ester [12].The exact function of SL-1 is still unknown but its absence in humans presents it as a potential target for anti-tubercular drug development [20].

The biosynthetic machinery of SL-1 involves seven important enzymes of which two are still unknown. These include Sulfotransferase (Stf0) that catalyses sulfation of trehalose to trehalose-2-sulfate (T2S), an acyltransferse, Polyketide synthase Associated Proteins A2 (PapA2) that adds straight chain palmitate or stearate to the 2-position of T2S yielding monoacyl SL659, another acyltransferse, Polyketide synthase Associated Proteins A1 (PapA1) that catalyses the transfer of the first phthioceranoyl or hydroxyphioceranoyl group to the 3-position of SL659 to create a diacyl SL1278, Polyketide synthase (Pks2) that transfers phthioceranoyl or hydroxyphioceranoyl group and the Large Membrane Protein (MmpL8) that transports SL1278 from the interior of the cell to exterior. There are two unidentified extracellular acyltransferases that catalyze the addition of the final two acyl units in SL-1 structure [30].

Pks2 is known to be involved in SL-1 assembly [24], thus deletion of Pks2 results in the loss of principle sulfolipid SL-1 in *Mycobacterium tuberculosis*. Of all the Pks genes linked to PAPs, Pks2 is only one located next to two Pap genes (PapA3 and Pks3/4 are involved in the biosynthesis of polyacylated trehalose, and the metabolite synthesized by PapA4 and Pks5 has not yet been identified) [33].

The sequencing of the *Mycobacterium tuberculosis* genome revealed five open reading frames, PapA1-PapA5, that were annotated as genes for polyketide synthase-associated proteins (PAPs) because of their linkage to their genes encoding enzymes involved in polyketide biosynthesis [7]. Most surface-exposed lipids are synthesized by specialized polyketide synthesis that elongates straight chain fatty acids by the stepwise addition of short acyl chains. Polyketide synthase-associated proteins, PapA2 and PapA1 reside within the same cluster of genes encoding Pks2, and MmpL8 in the *Mycobacterium tuberculosis* genome [4, 25, 34]. PapA1 is associated with Pks2, which encodes the enzyme that synthesizes the phthioceranic acids of SL-1. However, PapA2 is distal from Pks2 in the *Mycobacterium tuberculosis* genome and is not directly associated with another polyketide synthase [18].

Phthiocerol Dimycocerosate (PDIM) is the product of two such systems [1, 34]. An acyltransferase that installs the lipid groups on the *Mycobacterium tuberculosis* virulence factors phthiocerol Dimycocerosate (PDIM) was identified as the polyketide synthase associated proteins PapA5 [4, 25, 34]. This enzyme transfers a mycocerosyl group from the acyl carrier protein domain of the polyketide synthase [17].

Materials and methods

The proteins sequences and information (Name & Origin, Protein attributes, General annotation and Entry information) of PapA1 and PapA2 were retrieved from the Uniprot Knowledgebase (UniprotKB), in FASTA format. APSSP (Advanced Protein Secondary Structure Prediction Server) that predicted the secondary structures of proteins PapA1 and PapA2 from their amino acid sequences, by using nearest neighbor and neural network approach.

All the target protein sequences retrieved from Uniprot were BLASTed against Protein Data Bank Proteins (pdb) using the Protein-Protein BLAST (Psi-blast) program to find out the templates that carried more similarities with target proteins. Using the input files MODELLER 9v8 generated 3D model of target proteins.Work was done in the project mode by using DeepView (SwissPDB Viewer, or SPDBV 4.0.1) software that assured a better control of the homology modeling process and promoted a deeper understanding of various features of the protein structure you were going to model.

Friend, a bioinformatics application designed for simultaneous analysis and visualization of multiple structures and sequences of proteins and/or DNA/RNA. Energy calculations on a protein chain, were performed by ANOLEA (Atomic Non-Local Environment Assessment) and evaluated the "Non-Local Environment" (NLE) of each heavy atom in the molecule. A plot obtained from the Verify3D structure Evaluation Server that represented the average 3D-1D profile score.

Results

APSSP gave output that showed amino acid, secondary structure state (H->helix; E->Strand; C->Coil) and probability of correct prediction. In PapA1, 41.73% are H->helix, 15.49% are E->Strand, and 42.78% are C->Coil, whereas in PapA2, 45.47% are H->helix, 13.12% are E->Strand, and 41.40% are C->Coil.

Errors of the model are usually estimated either from the energy of the model or from the resemblance of a given characteristic of the model to real structures [32]. Predicted models of PapA1 and PapA2 proteins (Figs. 1-4) are evaluated by ANOLEA and Verify3D tools (Table 1).

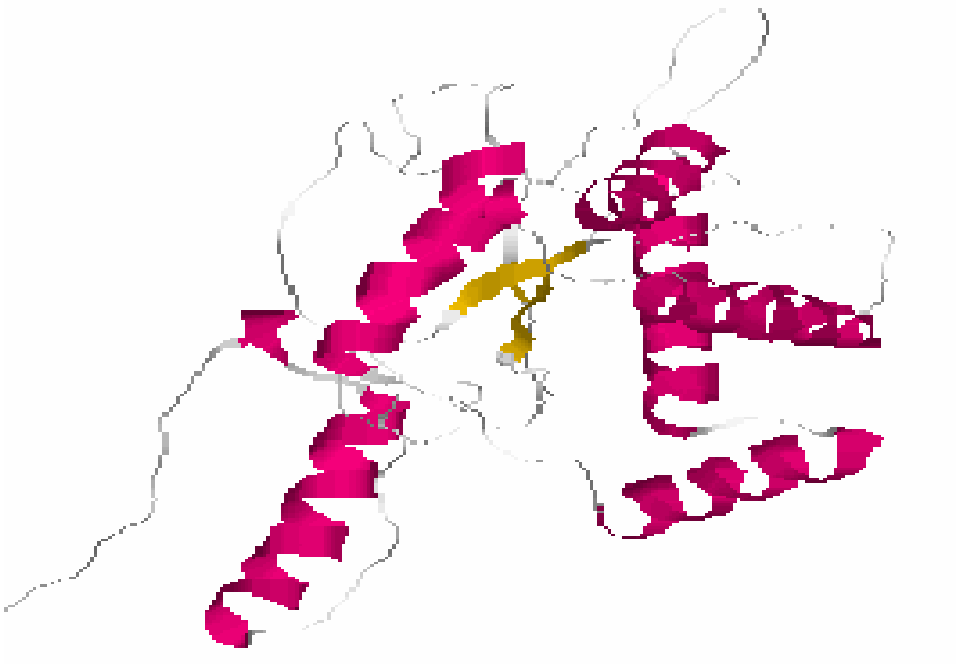

Fig. 1 Predicted model of PapA1 protein by Modeller program

Fig. 2 Predicted model of PapA2 protein by Modeller program

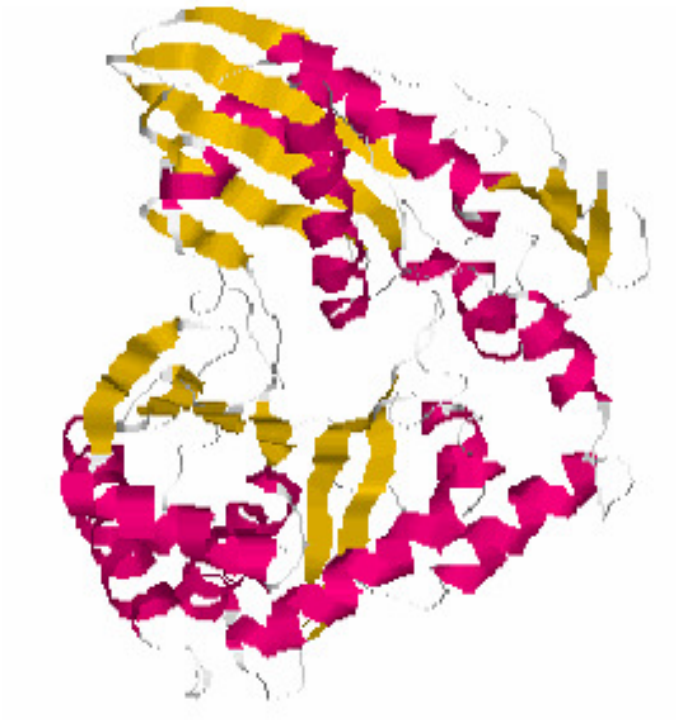

Fig. 3 Predicted model of PapA1 protein by Swiss-Model

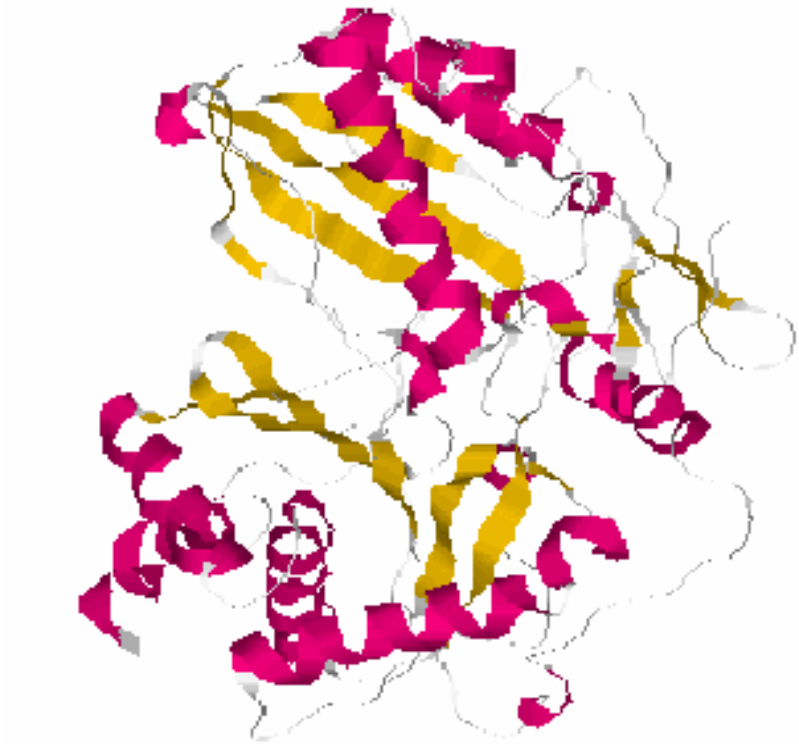

Fig. 4 Predicted model of PapA2 protein by Swiss-Model

Table 1. Models evaluations of proteins by using ANOLEA & Verify3D

Protein name	Modeling program	Evaluation program	High energy AA (%)	Z-score	3D-1D profile score
PapA1	Modeller	ANOLEA	68.72%	7.13	-
		Verify3D	-	-	0.51
	Swiss-Model	ANOLEA	71.39%	9.09	-
		Verify3D	-	-	0.73
PapA2	Modeller	ANOLEA	51.49%	5.00	-
		Verify3D	-	-	0.45
	Swiss-Model	ANOLEA	46.40%	4.25	-
		Verify3D	-	-	0.78

Discussion

In the present study, PapA1 and PapA2 are used for comparative analysis that involved in Sulfolipid (SL-1) biosynthesis pathway of *Mycobacterium tuberculosis*. The biosynthetic machinery of SL-1 involves seven important enzymes of which two are still unknown. MmpL8, Stf0, PapA1 and PapA2 were found to be very unique and entirely absent in human except Pks2 that showed significant alignment with human proteome. So that MmpL8, Stf0, PapA1, and PapA2 could be used as drug targets against *Mycobacterium tuberculosis* infection.

'3CIG' template of PapA1 and '2VSQ' template of PapA2, with 19% and 26% identities are selected. '3CIG' a Toll-Like Receptor (TLR3) is a member of TLR family that recognizes pathogen associated molecular signature and initiate inflammatory responses [11, 15, 16]. '2VSQ' Nonribosomal Peptide Synthases (NRPS) are modular multidomain enzymes that catalyze up to several dozen reactions in a highly coordinated manner to generate peptides of remarkable structural and functional diversity, and the NRPS assembly line logic is analogous of other megasynthases such as polyketide (PKS) and fatty acid synthases [35, 36]. It was very difficult to get maximum homology of these proteins with such a less number of PDB structures available.

Benchmark, of 3 different homology modeling programs – Modeller, Swiss-Model, and ESyPred3D, has been performed used to transform the alignment to a 3D model. The difference between the programs is how the information contained in the alignment is used to build a 3D model. One of the most frequently used modeling programs, Modeller generated 3D models by satisfaction of spatial restraints use a set of restraints derived from the alignment [30]. Swiss-Model assembled a model from a small number of rigid bodies obtained from the core of the aligned regions [3, 13]. The final three-dimensional (3D) structure was built using the modeling package Modeller [19]. 3D models of PapA1 and PapA2 were evaluated by evaluation tools, ANOLEA and Verify3D. Energy calculations are performed by ANOLEA that gave the Z-scores, 7.13 and 9.09, of PapA1 by Modeller and Swiss-Model respectively, indicating, that model obtained from modeler is better than Swiss-Model.

While the PapA2 models obtained by Modeller and Swiss-Model gave the 5.00 and 4.25 Z-scores. Plots obtained from the Verify3D structure Evaluation server that represented the average 3D-1D profile score. The 3D-1D profile score for the compatibility of the sequence with the model is the sum, over all residue positions, of the 3D-1D scores for the amino acid sequence of the protein [21]. 3D-1D profile scores, of PapA1 models are 0.51 and 0.73 and of PapA2 models are 0.45 and 0.75 obtained by Modeller and Swiss-Model respectively.

Two genes annotated as polyketide synthase-associated proteins, PapA2 and PapA1 reside within the same cluster of genes encoding Pks2, and MmpL8 in the *Mycobacterium tuberculosis* genome [4, 25, 34]. Genes involved in the synthesis and transport of these polyketides is required for bacterial growth and virulence in mice [6]. Two polyketide synthase associated acyltransferases are involved in the biosynthesis of the major sulfolipid-1 in *Mycobacterium tuberculosis* [2] and these could be used as drug targets and drug design against *Mycobacterium tuberculosis* infection. In future, structural conformations of these polyketide synthase-associated proteins, PapA1 and PapA2 could be used to find out the other proteins that not yet have been identified in this pathway.

References

1. Azad A. K., T. D. Sirakova., L. M. Rogers., P. E. Kolattukudy (1996). Targeted Replacement of the Mycocerosic Acid Synthase Gene in *Mycobacterium bovis* BCG Produces a Mutant that Lacks Mycosides, Proc. Natl. Acad. Sci. U.S.A., 93, 4787-4792.
2. Bhatt K., S. S. Gurcha, A. Bhatt, G. S. Besras, W. R. Jacobs (2007). Two Polyketide-synthase-associated Acyltransferases are Required for Sulfolipid Biosynthesis in *Mycobacterium tuberculosis*, Microbio., 153, 513-520.
3. Blundell T. L., B. L. Sibanda., M. J. Sternberg., J. M. Thornton (1987). Knowledge-based Prediction of Protein Structures and the Design of Novel Molecules, Nature, 326, 347-352.
4. Buglin J., K. C. Onwueme, J. A. Ferreras, L. E. Quadri, C. D. Lima (2004). Crystal Structure of PapA5, a Phthiocerol Dimycocerosyl Transferase from *Mycobacterium tuberculosis*, J. Biol. Chem., 279, 30634-30642.
5. Briken V., S. A. Porcelli, G. S. Besra, L. Kremer (2004). Mycobacterial Lipoarabinomannan and Related Lipoglycans: From Biogenesis to Modulation of the Immune Response, Mol. Microbiol., 53, 391-403.
6. Camacho L. R., D. Ensergueix, E. Perez, B. Gicquel, C. Guilhot (1999). Identification of a Virulence Gene Cluster of *Mycobacterium tuberculosis* by Signature-tagged Transposon Mutagenesis, Mol. Microbiol., 34, 257-267.
7. Cole S. T., R. Brosch, J. Parkhill, T. Garnier, C Churcher, D Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R Connor, R. Davies, K Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. Mclean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, B. G. Barrell (1998). Deciphering the Biology of *Mycobacterium tuberculosis* from the Complete Genome Sequence, Nature, 393, 537-544.
8. Converse S. E., J. D. Mougous, D. M. Leavell, J. A. Leary, C. R. Bertozzi, J. S. Cox (2003). MmpL8 is Required for Sulfolipid-1 Biosynthesis and *Mycobacterium tuberculosis* Virulence, Proc. Natl. Acad. Sci. U.S.A., 100(10), 6121-6126.
9. Domenech P., M. B. Reed, C. S. Dowd, C. Manca, G. Kaplan, C. E. Barry (2004). The Role of MmpL8 in Sulfatide Biogenesis and Virulence of *Mycobacterium tuberculosis*, J. Biol. Chem., 279(20), 21257-21265.
10. Gangadharam P. R., M. L. Cohn, G. Middlebrook (1963). Infectivity, Pathogenicity and Sulfolipid Fraction of Some Indian and British Stains of Tubercule bacilli, Tubercle, 44, 452-455.
11. Gay N. J., M. Gangloff (2007). Structure and Function of Toll Receptors and Their Ligands, Annu. Rev. Biochem., 76, 141-165.
12. Goren M. B., O. Brokl, B. C. Das (1976). Sulfatides of *Mycobacterium tuberculosis*: The Structure of the Principal Sulfatide (SL-I), Biochem., 15, 2728-2735.
13. Greer J. (1990). Comparative Modeling Methods: Application to the Family of the Mammalian Serine Proteases, Proteins, 7, 317-334.
14. Hardie R. M., J. M. Watson (1992). *Mycobacterium bovis* in England and Wales: Past, Present and Future, Epidemiol. Infect., 109(1), 23-33.
15. Iwasaki A., R. Medzhitov (2004). Toll-like Receptor Control of the Adaptive Immune Responses, Nat. Immunol., 5(10), 987-995.
16. Kaisho T., S. Akira (2006). Toll-like Receptor Function and Signaling, J. Allergy Clin. Immunol., 117(5), 979-987.

17. Kolattukudy P. E., N. D. Fernandes, A. K. Azad, A. M. Fitzmaurice, T. D. Sirakova (1997). Biochemistry and Molecular Genetics of Cell-wall Lipid Biosynthesis in Mycobacteria, Mol. Microbiol., 24(2), 263-270.
18. Kumar P., M. W. Schelle, M. Jain, F. L. Lin., C. J. Petzold, M. D. Leavell, J. A. Leary, J. S. Cox, C. R. Bertozzi (2007). PapA1 and PapA2 are Acyltransferases Essential for the Biosynthesis of the *Mycobacterium tuberculosis* Virulence Factor Sulfolipid-1, Proc. Natl. Acad. Sci. U.S.A., 104(27), 11221-11226.
19. Lambert C., N. Leonard, D. X. Bolle, E. Depiereux (2002). ESyPred3D: Prediction of Proteins 3D Structures, Bioinformatics, 18(9), 1250-1256.
20. Lin F. L., H. V. Halbeek, C. R. Bertozzi (2007). Synthesis of Mono and Dideoxygenated α,α-trehalose Analogs, Carbohydr. Res., 342(14), 2014-2030.
21. Luthy R., J. U. Bowie, D. Eisenberg (1992). Assessment of Protein Models with Three-dimensional Profiles, Nature, 356, 83-85.
22. Melo F., E. Feytmans (1997). Novel Knowledge-based Mean Force Potential at Atomic Level, Mol. Bio., 267, 207-222.
23. Middlebrook G., C. M. Coleman, W. B. Schaefer (1959). Sulfolipid from Virulent Tubercle bacilli, Proc. Natl. Acad. Sci. U.S.A., 45, 1801-1804.
24. Mougous J. D., C. J. Petzold, R. H. Senaratne, D. H. Lee, D. L. Akey, F. L. Lin, S. E. Munchel, M. R. Pratt, L. W. Riley, J. A. Leary, J. M. Berger, C. R. Bertozzi (2004). Identification, Function and Structure of the Mycobacterial Sulfotransferase that Initiates Sulfolipid-1 Biosynthesis, Nat. Struct. Mol. Biol., 11(8), 721-729.
25. Onwueme K. C., J. Ferreras, J. Buglino, C. D. Lima, L. E. Quadri (2004). Mycobacterial Polyketide-associated Proteins are Acyltransferases: Proof of Principle with *Mycobacterium tuberculosis* PapA5, Proc. Natl. Acad. Sci. U.S.A., 101, 4608-4613.
26. Petrey D., B. Honig (2003). GRASP2: Visualization, Surface Properties, and Electrostatics of Macromolecular Structures and Sequences, Meth. Enzymol., 374, 492-509.
27. Reed M. B., P. Domenech, C. Manca, H. Su, A. K. Barczak, B. N. Kreiswirth, G. Kaplan, C. E. Barry (2004). A Glycolipid of Hypervirulent Tuberculosis Strains that Inhibits the Innate Immune Response, Nature, 431, 84-87.
28. Rousseau C., N. Winter, E. Pivert, Y. Bordat, O. Neyrolles, P. Ave, M. Huerre, B. Gicquel, M. Jackson (2004). Production of Phthiocerol Dimycocerosates Protects *Mycobacterium tuberculosis* from the Cidal Activity of Reactive Nitrogen Intermediates Produced by Macrophages and Modulates the Early Immune Response to Infection, Cell Microbiol., 6, 277-287.
29. Ryll R., K. Watanabe, N. Fujiwara, H. Takimoto, R. Hasunuma, Y. Kumazawa, M. Okada, I. Yano (2001). Mycobacterial Cord Factor, but not Sulfolipid, Causes Depletion of NKT Cells and Upregulation of CD1d1 on Murine Macrophages, Microbes Infect., 3(8), 611-619.
30. Sali A., T. L. Blundell (1993). Comparative Modelling by Statisfaction of Spatial Restraints, J. Mol. Biol., 234, 779-815.
31. Schelle M. W., C. R. Bertozzi (2006). Sulfate Metabolism in Mycobacteria, Chem. Bio. Chem., 7, 1516-1524.
32. Sippl M. J. (1995). Knowledge-based Potentials for Proteins, Curr. Opin. Struct. Biol., 5(2), 229-235.
33. Sirakova T. D., A. K. Thirumala, V. S. Dubey, H. Sprecher, P. E. Kolattukudy (2001). The Mycobacterium Tuberculosis pks2 Gene Encodes the Synthase for the Hepta- and Octamethyl-branched Fatty Acids Required for Sulfolipid Synthesis, J. Biol. Chem., 276(20), 16833-16839.

34. Trivedi O. A., P. Arora, A. Vats, M. Z. Ansari, R. Tickoo, V. Sridharan, D. Mohanty, R. S. Gokhale (2005). Dissecting the Mechanism and Assembly of a Complex Virulence Mycobacterial Lipid, Mol. Cell., 17, 631-643.
35. Walsh C. T. (2004). Polyketide and Nonribosomal Peptide Antibiotics: Modularity and Versatility, Science, 303(5665), 1805-1810.
36. Weissman K. J., R. Muller (2008). Crystal Structure of a Molecular Assembly Line, Ange. Chem. Inter. Ed., 47(44), 8344-8346.
37. WHO (2008). Global Tuberculosis Control: Profiles of High-burden Countries. W.H.O. (http://www.who.int/tb/publications/global_report/2008/pdf/annex_1.pdf).
38. Xiong J. (2006). Protein Tertiary Structure Prediction, In: Essential Bioinformatics, Cambridge University Press, New York, 214-230.

Rana Adnan Tahir, B.Sc.
E-mail: ranaadnantahir@yahoo.com

Rana Adnan Tahir s/o Tahir Mahmood is an active member of Research Group at International Islamic University Islamabad Pakistan. He is currently the student of M.Sc. Bioinformatics in International Islamic University Islamabad Pakistan. He has passed his B.Sc. (Hons) in Bioinformatics from Government College University Faisalabad Pakistan in 2010. In research group, he is working on ongoing research projects. He has command over many Bioinformatics data analysis and structure prediction tools. His interests are Structural Bioinformatics, Phylogenetic Analysis, Docking, Comparative Genomics, and Data Mining.

Sheikh Arslan Sehgal, B.Sc.
E-mail: arslansehgal@yahoo.com

Sheikh Arslan Sehgal s/o Sheikh Muhammad Ashraf Advocate is a dedicated member of Research Group at International Islamic University Islamabad Pakistan. He is currently the student of M.Sc. Bioinformatics in International Islamic University Islamabad Pakistan. He graduated from Government College University in 2010. In Research group, he works as a database Developer and is working as an active participant in ongoing research projects. He has command over many Bioinformatics tools and software. He also works on various programming languages like Java, Perl and Oracle. His core interests are Phylogenetic analysis, Biological Database development, Drug discovery and Computer Aided Drug Designing.

Ambreen Ijaz, M.Sc.
E-mail: ambreen_7868@yahoo.com

Ambreen Ijaz D/o Ijaz Akbar Khan is working as lecturer in the Department of Bioinformatics and Biotechnology GC University Faisalabad. She did her M.Sc. in Chemistry from University of Agriculture, Faisalabad, Pakistan and M.Phil. in Biotechnology from Quaid-i-Azam University, Islamabad, Pakistan. She has specialization in the field of Plant Biotechnology and Microbiology and worked as a Researcher in National Institute for Biotechnology and Genetic Engineering Faisalabad, Pakistan. She is currently working on projects of Structure Prediction, Comparative Genomics and Microbiology.