Interdiscip Sci Comput Life Sci (2014) 6: 48–56
DOI: 10.1007/s12539-014-0188-y

Application of a Subtractive Genomics Approach for *in silico* Identification and Characterization of Novel Drug Targets in *Mycobacterium tuberculosis* F11

Md. Ismail HOSEN[1,2*], Arif Mohammad TANMOY[1], Deena-Al MAHBUBA[1], Umme SALMA[1], Mohammad NAZIM[3], Md. Tariqul ISLAM[1], Sharif AKHTERUZZAMAN[1]
[1](Department of Biochemistry and Molecular Biology, Faculty of Biological Science, University of Dhaka, Dhaka 1000, Bangladesh)
[2](German Cancer Research Center, Heidelberg 69120, Germany)
[3](Nagoya University Graduate School of Medicine, Nagoya University, Nagoya 4640856, Japan)

Received 21 October 2012 / Revised 27 November 2012 / Accepted 4 December 2012

Abstract: Extensive dead ends or host toxicity of the conventional approaches of drug development can be avoided by applying the *in silico* subtractive genomics approach in the designing of potential drug target against bacterial diseases. This study utilizes the advanced *in silico* genome subtraction methodology to design potential and pathogen specific drug targets against *Mycobacterium tuberculosis*, causal agent of deadly tuberculosis. The whole proteome of *Mycobacterium tuberculosis* F11 containing 3941 proteins have been analyzed through a series of subtraction methodologies to remove paralogous proteins and proteins that show extensive homology with human. The subsequent exclusion of these proteins ensured the absence of host cytotoxicity and cross reaction in the identified drug targets. The high stringency (expectation value 10^{-100}) analysis of the remaining 2935 proteins against database of essential genes resulted in 274 proteins to be essential for *Mycobacterium tuberculosis* F11. Comparative analysis of the metabolic pathways of human and *Mycobacterium tuberculosis* F11 by KAAS at the KEGG server sorted out 20 unique metabolic pathways in *Mycobacterium tuberculosis* F11 that involve the participation of 30 essential proteins. Subcellular localization analysis of these 30 essential proteins revealed 7 proteins with outer membrane potentialities. All these proteins can be used as a potential therapeutic target against *Mycobacterium tuberculosis* F11 infection. 66 of the 274 essential proteins were uncharacterized (described as hypothetical) and functional classification of these proteins showed that they belonged to a wide variety of protein classes including zinc binding proteins, transferases, transmembrane proteins, other metal ion binding proteins, oxidoreductase, and primary active transporters *etc.* 2D and 3D structures of these 15 membrane associated proteins were predicted using PRED-TMBB and homology modeling by Swiss model workspace respectively. The identified drug targets are expected to be of great potential for designing novel anti-tuberculosis drugs and further screening of the compounds against these newly targets may result in discovery of novel therapeutic compounds that can be effective against *Mycobacterium tuberculosis*.
Key words: essential proteins, membrane associated proteins, DEG, tuberculosis, metabolic pathways and protein-protein interaction.

1 Introduction

The current study makes use of the subtractive genomics approach, database of essential gene (DEG), differential pathway analysis and sub-cellular localization prediction to analyze the complete proteome of *Mycobacterium tuberculosis* F11 to search for potential drug and vaccine targets against tuberculosis.

Tuberculosis or TB is a common, and in many cases lethal infectious disease caused by various strains of Mycobacterium, usually *Mycobacterium tuberculosis* (MTB) (Kumar, 2007). One third of the world's population is thought to be infected with *M. tuberculosis* (Dolin *et al.*, 1994; Jasmer *et al.*, 2002; WHO, 2010) and new infections occur at a rate of about one per second (WHO, 2010). In 2010 about 1.45 million people died, mostly in developing countries (WHO, 2011). In addition, a number of low-income and middle-income countries account for more than 80% of the active cases of tuberculosis in the world. Due to the devastating effect of HIV on susceptibility to tuberculosis, sub-Saharan Africa has been dis-

*Corresponding author.
E-mail: milon_ismail_02@yahoo.com

proportionately affected and accounts for four of every five cases of HIV-associated tuberculosis (Lawn and Zumla, 2011). Although there are available treatment approaches against TB, due to the emergence of totally drug resistant tuberculosis (Zarir *et al.*, 2011), efforts are continued to design new targets. Designing of potential drug targets against TB can thus mitigate the problems associated with this disease.

Out of the >1000 bacterial genomes sequenced to date, most are of pathogenic bacteria that cause diseases. All these projects of genome sequencing have generated a huge amount of analyzable data that can be fully exploited for the identification and characterization of virulent factors in these pathogens, to identify novel putative targets for therapeutic intervention (Miese *et al.*, 2003).

Bioinformatics-based methodologies provide alternative approaches and have the potential to cut costs in identifying drug targets, as conventional laboratory based experiments to identify candidate molecules as drug targets are expensive and require a huge amount of time. Whole proteome of host and pathogen can be utilized by a subtractive genomics approach to identify proteins exclusively present in the pathogen by deducing the homologous proteins. This speedy procedure has been used successfully to identify novel drug targets in pathogens like *Edwardsiella tarda* (Neema *et al.*, 2011), *Pseudomonas aeruginosa* (Sakharkar *et al.*, 2004), *Helicobacter pylori* (Dutta *et al.*, 2006), *Salmonella typhi* (Rathi *et al.*, 2009) and *Neisseria meningitides* (Sarangi *et al.*, 2009). In this study we employed a technique that makes use of the systems of the Cluster Database at High Identity with Tolerance (CD-HIT), Basic Local Alignment Search Tool for Proteins (BLASTP), Database of Essential Genes (DEG), KEGG Automatic Annotation Server (KAAS), Kyoto Encyclopedia of Genes and Genomes (KEGG), PSORTb, and SWISS-MODEL to identify, characterize, analyze and predict the structures of the essential genes of the human pathogen *Mycobacterium tuberculosis* that could be suitable vaccine or drug candidates.

Here, in this study we have used *Mycobacterium tuberculosis* F11 strains as our target pathogen. This F11 strain of *Mycobacterium tuberculosis* is the most successful strain of *M. tuberculosis* identified in the Western Cape communities in South Africa (Warren *et al.*, 1999; Warren *et al.*, 2002) and also has a global perspective (Victor*et al.*, 2004).

We demonstrated the unprecedented potential of the available computational databases, and the application of a subtractive genomics approach for the identification of essential genes that may be considered as candidates for antibacterial drug discovery, using the completely sequenced *Mycobacterium tuberculosis* F11 genome as an example. Furthermore, our approach successfully identified a number of promising protein targets for the development of new therapeutic targets.

2 Materials and methods

The flow chart describing the detailed methodology for identifying pathogen specific essential proteins in *Mycobacterium tuberculosis* F11 for characterizing potential drug targets is shown in Fig. 1.

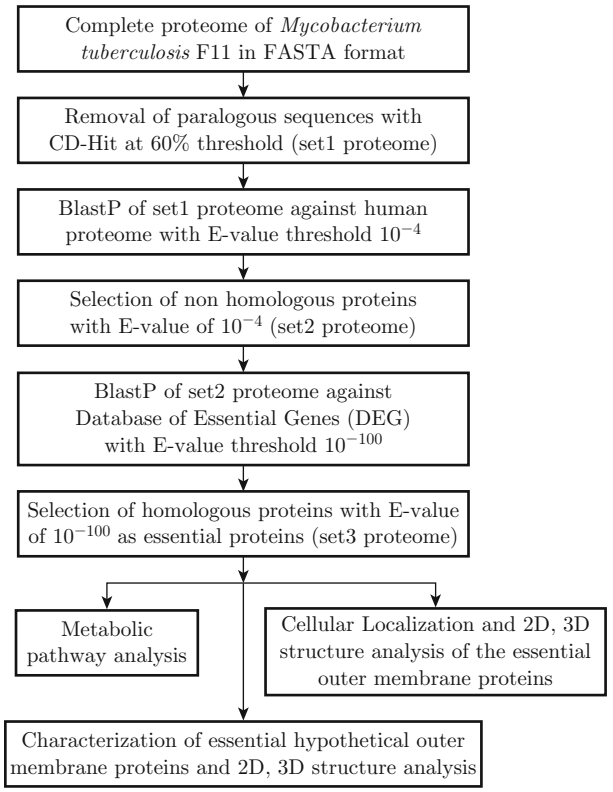


Fig. 1 Flowchart for identification of potential drug targets in *M. tuberculosis* F11.

2.1 Retrieval of proteomes of pathogen and host

The complete proteome of *Mycobacterium tuberculosis* F11 was retrieved from SwissProt (Bairoch and Apweiler, 2000) and complete *Homo sapiens* proteome was downloaded from NCBI (Maglott *et al.*, 2007). Prokaryotic essential protein sequences were retrieved from the database of essential genes (DEG) (Zhang *et al.*, 2004) (http://tubic.tju.edu.cn/deg/).

2.2 Identification of essential proteins in *Mycobacterium tuberculosis*

The whole proteome of *Mycobacterium tuberculosis* F11 was purged at 60% threshold using CD-HIT (Li and Godzik, 2006) to exclude paralogs or duplicates for further analysis. The set of proteins obtained after excluding the paralogs were subject to BlastP against *Homo sapiens* proteome with the expectation value (E-value) cut-off of 10^{-4}. The

resultant dataset were with no homologs in *Homo sapiens*. BlastP analysis was performed for the non homologous protein sequences of *M. tuberculosis* against DEG with E-value cut-off score of 10^{-100}. A minimum bit score cut-off of 100 was used to screen out genes that appeared to represent essential genes. The protein sequences obtained are non-homologous essential proteins of *M. tuberculosis*. Stand-alone BLAST software (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST/blast-2.2.21-ia32-win32.exe) was used for the BlastP analysis. E-value cut-off thresholds for BlastP and DEG search were set by using the command line parameter option in standalone BLAST.

2.3 Functional classification of the uncharacterized essential proteins

Functional family prediction of the putative uncharacterized (designated also as hypothetical) essential proteins was done by using the SVMProt web server (http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi) (Cai *et al.*, 2003). SVMProt utilizes Support Vector Machine for classification of a protein into functional family from its primary sequence.

2.4 Metabolic pathway analysis

Metabolic pathway analysis of the essential proteins of *M. tuberculosis* was done by KAAS (KEGG Automatic Annotation Server) at KEGG for the identification of potential targets. KAAS provides functional annotation of genes by BLAST comparisons against the manually curated KEGG GENES database. The result contains KO (KEGG Orthology) assignments and automatically generated KEGG pathways (Moriya *et al.*, 2007).

2.5 Sub-cellular localization prediction

Protein sub-cellular localization prediction involves the computational prediction of where a protein resides in a cell. Prediction of protein sub-cellular localization is an important component as it predicts the protein function and genome annotation, and it can aid the identification of targets. Computational tools are involved in the prediction of sub-cellular localization, which illustrates where the protein resides in the cell. So if the protein is present in the outer membrane, there is more possibility that it can be highlighted as a potential drug target. The program PSORTb V.3.0 (http://www.psort.org/psortb/index.html) was used for sub-cellular localization prediction (Gardy *et al.*, 2005).

2.6 Protein-protein interaction prediction of the hypothetical essential outer membrane proteins

The hypothetical proteins that were found to be essential and have the outer membrane properties as predicted by the database of essential genes (DEG) and PSORTb respectively were predicted for their interactions with other proteins. STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources: genomic context, high-throughput experiments, (conserved) co-expression and previous knowledge. STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently covers 5,214,234 proteins from 1133 organisms (Szklarczyk, 2011). This tool was used to study the protein-protein interactions of the hypothetical essential outer membrane proteins.

2.7 Two-dimensional (2D) and three-dimensional (3D) structure prediction of the target proteins

To obtain a 3D structure of the OMPs, homology modeling of the outer membrane was done by the server SWISS-MODEL (swissmodel.expasy.org). The server builds the structure based on the homology on the alignment between the target and template (Arnold *et al.*, 2006). 2D structure prediction of OMPs was done by PRED-TMBB (http://biophysics.biol.uoa.gr/PRED-TMBB/). The method predicts the transmembrane beta-strands of OMP (Bagos *et al.*, 2004).

3 Results and discussions

The result of this study explains the use of an elegant subtractive genomics approach to identify and characterize potential drug targets against *Mycobacterium tuberculosis* which is the causative agent of deadly tuberculosis. The *in silico* analysis shows that the whole proteome of *Mycobacterium tuberculosis* F11 contains a total of 3941 proteins (See Table 1 for a summary of the number of proteins after each step of analysis). A total of 213 proteins out of these 3941 proteins are redundant proteins as found and sorted out by the CD-HIT tools. The remaining 3728 proteins are non-redundant *Mycobacterium tuberculosis* F11 proteins. These proteins were then used for a BlastP analysis against the human proteome to select the proteins that are non-homologous to human. By sorting out the BlastP results with a threshold E-value of 10^{-4}, we got 793 proteins to show homology with human proteins and these proteins were avoided in further analysis as their homology with the host (human, in this case) proteins can lead to unwanted toxicity and cross-reactivity in human if they are selected as a drug target against tuberculosis. That's why the non-homologous proteins

were selected for further identification of the essential genes of *Mycobacterium tuberculosis* by BlastP search in DEG that has a collection of essential genes among a wide range of pathogenic and non-pathogenic organism (both pro- and eukaryotes). As the essential genes are often the favorable targets for antibiotics because of the involvement of essential genes in essential cellular processes in pathogenic bacteria, they can be suitable and promising new targets for designing antibacterial drugs. We were very stringent in sorting out essential proteins in *Mycobacterium tuberculosis* by setting out a threshold E-value of 10^{-100}.

Table 1 Subtractive proteomic and metabolic pathway analysis result for *Mycobacterium tuberculosis* F11

Features of *M. tuberculosis* F11	Number
Total number of proteins	3941
Duplicates (>60% identity) in CD-HIT	213
Non-paralogs	3728
Non-human homologous proteins (E-value 10^{-4})	2935
Essential protein in DEG (E-value 10^{-100})	274
Essential proteins involved in metabolic pathways	168
Pathways unique to the organism (*M. tuberculosis*)	20
Proteins involved in unique pathways	30
Membrane associated non-human homolog of essential genes	7

A total of 274 proteins were sorted as essential proteins and among these 66 were hypothetical uncharacterized proteins whose functional analysis by SVM-Prot shows that they are involved in multiple functioning, from zinc binding (23 proteins) to transmembrane potentialities (7 proteins) (Table 2).

The list of potential drug targets encoded in microbial genomes includes genes involved in peptidoglycan and cell wall synthesizing proteins, DNA replication, repair, transcription, translation, outer-membrane proteins, enzymes of intermediary metabolism, host-interaction factors *etc.*

As anticipated, our analysis found that many of the essential genes were involved in different metabolic pathway as analyzed by KEGG-KAAS. The comparative analysis of the metabolic pathway of human and *Mycobacterium tuberculosis* revealed that the essential proteins of *Mycobacterium tuberculosis* were involved in 20 unique pathogen specific metabolic pathways. A total of 30 essential proteins of *M. tuberculosis* are involved in these unique metabolic pathways which can be used as novel therapeutic targets. The 30 essential proteins are involved in the following unique metabolic pathways: C5-branched dibasic acid metabolism (2), carbon fixation pathways in prokaryotes (1), methane metabolism (2), D-alanine metabolism (2), lipopolysaccharide biosynthesis (1), peptidoglycan biosynthesis (2), geraniol degradation (1), polyketide sugar unit biosynthesis (1), biosynthesis of siderophore group nonribosomal peptides (2), phenylpropanoid biosynthesis (1), tropane, piperidine and pyridine alkaloid biosynthesis (1), streptomycin biosynthesis (1), novobiocin biosynthesis (1), nitrotoluene degradation (1), atrazine degradation (1), bacterial secretion system (3), two-component system (4), cell cycle - caulobacter (2), vibrio cholerae pathogenic cycle (1).

Table 2 Functional classification of putative uncharacterized essential proteins by SVM-Prot web server

Functional classification by SVMProt	Number of putative uncharacterized proteins
Transferases	18
Transmembrane	7
Outer membrane	1
Zinc-binding	23
All lipid-binding proteins	1
Hydrolases	1
Iron-binding	7
Major facilitator family	1
DNA replication	1
Lyases	1
Copper-binding	1
Lipoprotein	1
Oxidoreductases	1
Primary Active Transporters	1
Manganese-binding	1

The peptidoglycan layer is important for cell wall structural integrity, especially in Gram-positive organisms and also plays an important role in pathogenesis as it provides resistance to osmotic lysis. As D-alanine is the central molecule in the assembly and cross-linking of peptidoglycan, any protein involved in this process can be a very good drug target. We found that alanine racemase (EC:5.1.1.1) involved in D-alanine metabolism (conversion of L-alanine to D-alanine) was an essential protein of *Mycobacterium tuberculosis* that is also involved in the unique pathogen specific metabolic pathway. Our *in silico* analysis also found D-alanyl-alanine synthetase A (EC:6.3.2.4) (ddlA) as an important target as it is also involved in D-alanine metabolism that leads to the biosynthesis of peptidoglycan. Both of these proteins can be a potential target against *Mycobacterium tuberculosis* as they are essential in this organism and are absent in the host human.

Lipopolysaccharides (LPS) are also one of the main constituents of the outer cell wall of Gram-negative bacteria and play an important role for the survival of the pathogen. The enzyme lipid A biosynthesis lauroyl acyltransferase [EC:2.3.1.-] is an important enzyme in-

volved in the lipopolysaccharide biosynthesis pathway (acylates the intermediate (KDO)2-lipid IVA to form (KDO)2-(lauroyl)-lipid IVA). We found this enzyme as unique in the sense that it does not have any homology with the human proteins and hence can be a suitable drug target enhancing LPS specific defence against tuberculosis.

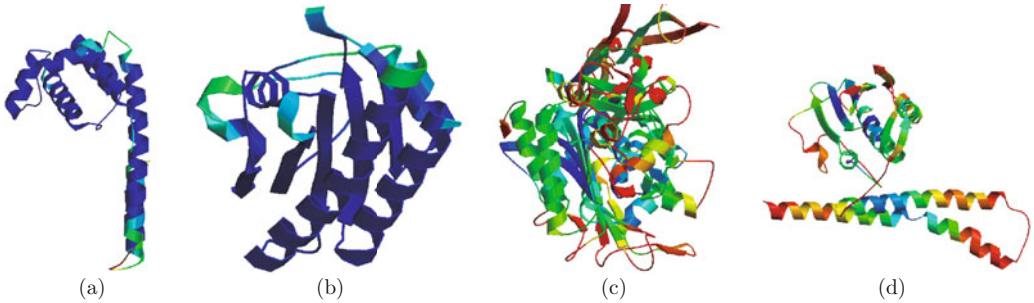
Two-component signal transduction systems enable bacteria to sense, respond, and adapt to any kind of stress situations by regulating gene expression. Subtractive genomics approach enabled us to pin-point four pathogen specific essential proteins (by DEG and KEGG analysis) in *Mycobacterium tuberculosis* that are involved in two-component signal transduction systems. Out of these four proteins two-component system, OmpR family, sensor histidine kinase MtrB [EC: 2.7.13.3] and two-component system, NarL family, sensor histidine kinase DevS [EC: 2.7.13.3] are involved in temperature and hypoxia induced stress response pathway respectively. The third essential protein, transcriptional regulatory protein kdpE is involved in the response to low turgor pressure as found in the KEGG pathway. Chromosomal replication initiation protein is the fourth and final essential protein in this system. All these four proteins are potentially important drug target as they can be targeted to hit the bacterial stress responsive pathways and hence can put pressure in the bacteria.

In Gram-positive bacteria, secreted proteins are commonly translocated across the single membrane by the general Secretory (Sec) pathway or the two-arginine targeting (Maglott *et al.*, 2007) pathway. We found three essential proteins involved in the unique pathogen specific pathways of *Mycobacbacterium tuberculosis*. Among these, Sec-independent protein translocase protein TatC is involved in the Tat pathway. Putative inner membrane protein translocase component YidC and preprotein translocase subunit SecA are involved in the sec-SRP pathway. These three essential proteins can also be potential drug targets.

Proteins with outer membrane potentiality have important role in the interaction with hosts in bacterial pathogenicity, playing a role in adherence, uptake of nutrients from the host, and countering host defense mechanisms (Seltmann, 2002). They could be protective antigens because the components of the outer membrane are easily recognized as foreign substances by immunological defense systems of hosts. The essential proteins involved in the unique metabolic pathways of *M. tuberculosis* F11 were therefore analyzed for their localization in the outer membrane. The program PSORTb identified seven proteins that reside in the outer membrane (Table 3). Fig. 2 shows the homology modeled structure of the OMPs of *M. tuberculosis*. Because Protein Data Bank structures homologous to the proteins YP_001289883.1, YP_001288061.1 and YP_001288119.1 were not available, the 2D structure prediction of these proteins was carried out using PRED-TMBB (Figs. 3(a)-3(c)), which predicts the number of transmembrane beta-barrels in the outer membrane. The results show that the proteins identified as OMPs by PSORTb had beta-barrels characteristic of such proteins, thus supporting the prediction.

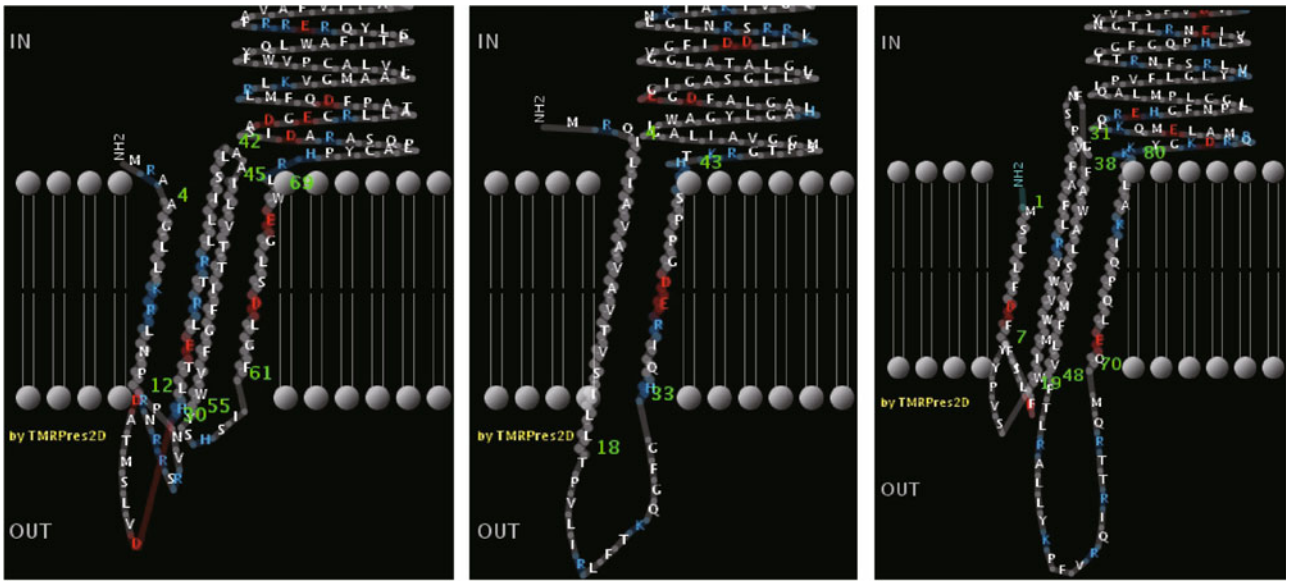
Table 3 List of the outer membrane proteins of *Mycobacterium tuberculosis* F11 identified by PSORTb

Accession No.	Name of protein	Subcellular localization
YP_001288126.1	Penicillin-binding membrane protein pbpB	Outer membrane
YP_001289883.1	Putative inner membrane protein translocase component YidC	Outer membrane
YP_001288061.1	Sec-independent protein translocase transmembrane protein tatC	Outer membrane
YP_001289201.1	Two component system sensor transduction histidine kinase mtrB	Outer membrane
YP_001289080.1	Two component system sensor histidine kinase devS	Outer membrane
YP_001286002.1	Replicative DNA helicase	Outer membrane
YP_001288119.1	Phospho-N-acetylmuramoyl-pentapeptide-transferase	Outer membrane



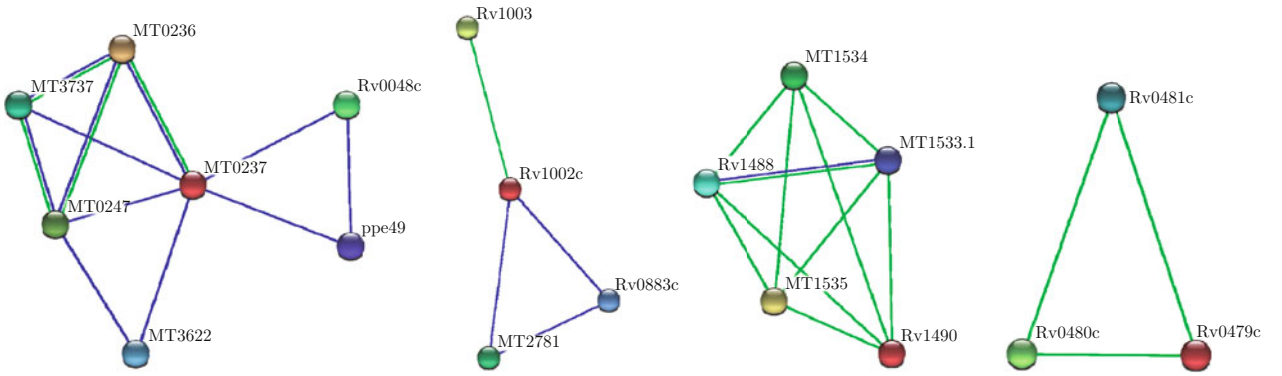
(a) (b) (c) (d)

Fig. 2 Homology modeled structure of the OMPs of *M. tuberculosis*. (a) YP_001286002.1, (b) YP_001288126.1, (c) YP_001288126.1, (d) YP_001289201.1.



(a) (b) (c)

Fig. 3 2D structures of (a) YP_001289883.1, (b) YP_001288061.1 and (c) YP_001288119.1 using PRED-TMBB.



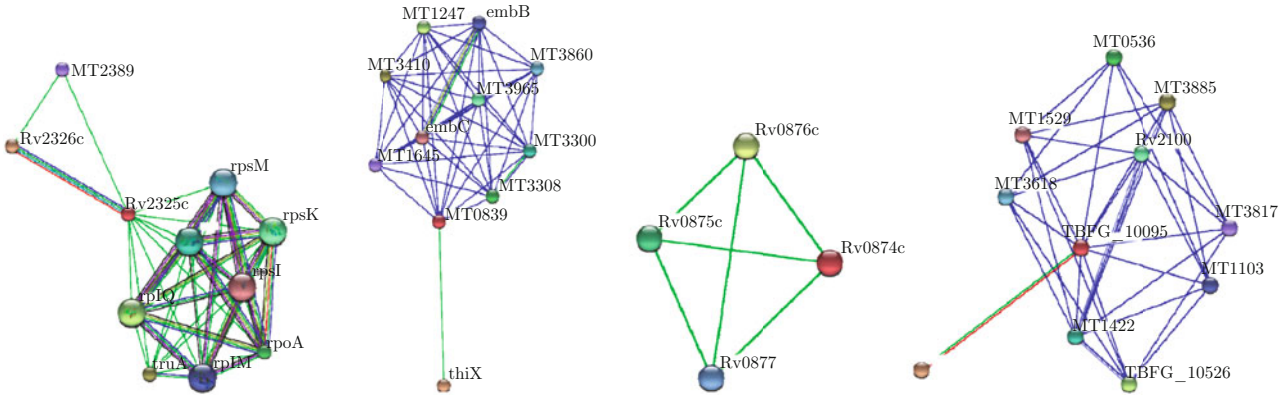
>gi|148821420|ref|YP_001286174.1| hypothetical protein (MT0237)

>gi|148822212|ref|YP_001286966.1| hypothetical protein (Rv1002c)

>gi|148822711|ref|YP_001287465.1| hypothetical protein (Rv1490)

>gi|148821678|ref|YP_001286432.1| hypothetical protein (Rv0479c)

(a)



>gi|148823525|ref|YP_001288279.1| hypothetical protein (Rv2325c)

>gi|148822023|ref|YP_001286777.1| hypothetical protein (MT0839)

>gi|148822082|ref|YP_001286836.1| hypothetical protein (Rv0874c)

>gi|148821286|ref|YP_001286040.1| hypothetical protein (TBFG_10095)

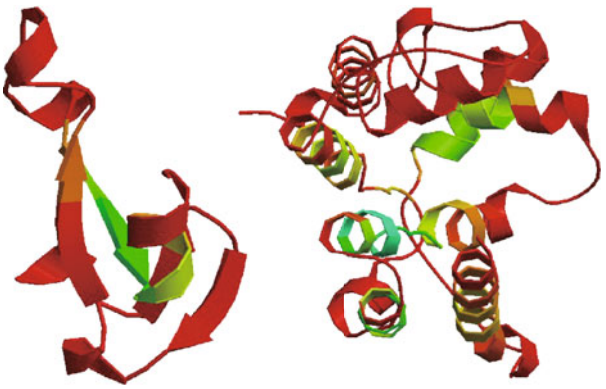
(b)

Fig. 4 Protein-protein interactions of hypothetical essential proteins with outer membrane properties predicted by STRING tool.

PSORTb analysis of the 66 putative uncharacterized essential proteins showed that 8 proteins are of outer membrane potentialities. We used STRING tool to predict the protein-protein interactions of the eight hypothetical essential proteins with outer membrane properties (Figs. 4(a) and 4(b)).

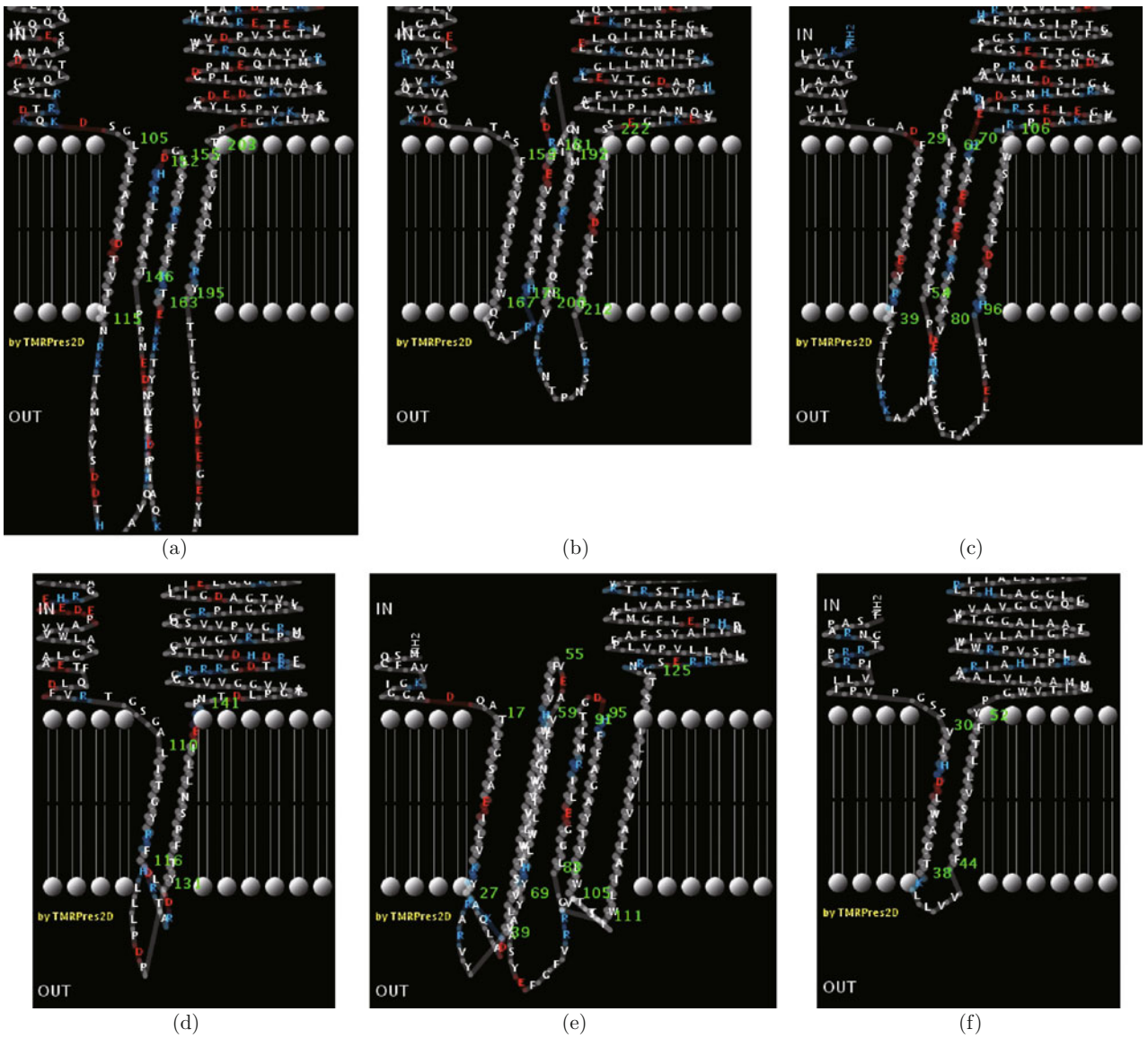
The 3D structures of these essential proteins with outer membrane properties were assessed by Swiss modeling. However, the Protein Data Bank structure homologous to 6 of these proteins was not available and the 2D structure prediction of these proteins was carried out using PRED-TMBB. Results are shown in Fig. 5 and Fig. 6.

As we mentioned in earlier section, our study resulted in a number of proteins that were hypothetical (or still



(a) (b)

Fig. 5 3D structures of the uncharacterized essential proteins. (a) YP_001286040.1, (b) YP_001286966.1.



(a) (b) (c)

(d) (e) (f)

Fig. 6 2D structures of the uncharacterized essential proteins. (a) YP_001286174.1, (b) YP_001286432.1, (c) YP_001286777.1, (d) YP_001286836.1, (e) YP_001287465.1, and (f) YP_001288279.1.

uncharacterized), but are essential for *Mycobacterium tuberculosis* F11. The essential hypothetical proteins with outer membrane properties provoked us to look for their interactions with other proteins. Using the STRING tool, we predicted these interactions and the results provided us with a good insight of these hypothetical proteins.

Protein-protein interaction study of the essential hypothetical proteins with outer membrane properties showed that these proteins interacted with a wide array of proteins including various transmembrane proteins, ABC transporter ATP-binding protein, tRNA pseudouridine synthase A, DNA-directed RNA polymerase subunit alpha, 30S ribosomal protein S11, thioredoxin ThiX, integral membrane indolylacetyl inositol arabinosyltransferase *etc.* The prediction confidences as expressed by scores are also provided in this table.

Some hypothetical proteins also interact with PPE family proteins that have been shown to be related to antigenic variation of *Mycobacterium tuberculosis* (Cole ST 1998). All these interactions revealed that we can propose these proteins as potential drug targets as well along with the seven essential outer membrane proteins that are characterized.

Finally we ended up with 15 proteins that are potential target proteins for drug designing against tuberculosis. These proteins need to be characterized for their drug target properties *in vitro* in future to achieve newer generation anti-tuberculosis drugs. In future, we are also planning to use these targets to carry out a laboratory based experiments to uncover novel therapeutic compounds.

4 Concluding remarks

Obstacles like toxicity or dead ends that are encountered in classical approaches of drug designing can be overcome by the use of the subtractive genomics approach that can speed up drug discovery in newer dimensions. Presumably, the candidate drug targets, which we have sorted out from *Mycobacterium tuberculosis* whole proteome, will accelerate the discovery of novel therapeutic agents against tuberculosis.

Acknowledgements

We cordially thank Prof. M. Anwar Hossain, Ms. Taslima Haque and Mr. Sajib Chakrabarty of the Department of Biochemistry and Molecular Biology, University of Dhaka, for their suggestions and inspiration during our research proceedings.

References

[1] Arnold, K., Bordoli, L., Kopp, J., Schwede, T. 2006. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. Bioinformatics 22, 195-201.

[2] Bagos, P.G., Liakopoulos, T.D., Spyropoulos, I.C., Hamodrakas, S.J. 2004. PRED-TMBB: A web server for predicting the topology of beta-barrel outer membrane proteins. Nucl Acid Res 32, W400-W404.

[3] Bairoch, A., Apweiler, R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucl Acid Res 28, 45-48.

[4] Cai, C.Z., Han, L.Y., Ji, Z.L., Chen, X., Chen, Y.Z. 2003. SVM-Prot: Web-based support vector machine software for functional classification of a protein from its primary sequence. Nucl Acid Res 31, 3692-3697.

[5] Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E. 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., Barrell, B.G. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393, 537-544.

[6] Dolin, P.J., Raviglione, M.C., Kochi, A. 1994. Global tuberculosis incidence and mortality during 1990-2000. Bull World Health Organ 72, 213-220.

[7] Dutta, A., Singh, S.K., Ghosh, P., Mukherjee, R., Mitter, S., Bandyopadhyay, D. 2006. *In silico* identification of potential therapeutic targets in the human pathogen Helicobacter pylori. In Silico Biol 6, 43-47.

[8] Gardy, J.L., Laird, M.R., Chen, F., Rey, S., Walsh, C.J., Ester, M., Brinkman, F.S. 2005. PSORTb v.2.0: Expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. Bioinformatics 21, 617-623.

[9] Jasmer, R.M., Nahid, P., Hopewell, P.C. 2002. Clinical practice. Latent tuberculosis infection. N Engl J Med 347, 1860-1866.

[10] Kumar, V., Abbas, A.K., Fausto, N., Mitchell, R.N. 2007. Robbins Basic Pathology. Saunders Elsevier, Philadelphia.

[11] Lawn, S.D., Zumla, A.I. 2011. Tuberculosis. Lancet 378, 57-72.

[12] Li, W., Godzik, A. 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658-1659.

[13] Maglott, D., Ostell, J., Pruitt, K.D., Tatusova, T. 2007. Entrez Gene: Gene-centered information at NCBI. Nucl Acid Res 35, D26-D31.

[14] Miesel, L., Greene, J., Black, T.A. 2003. Genetic strategies for antibacterial drug discovery. Nat Rev Genet 4, 442-456.

[15] Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., Kanehisa, M. 2007. KAAS: An automatic genome annotation and pathway reconstruction server. Nucl Acid Res 35, W182-W185.

[16] Neema, M., Karunasagar, I. 2011. *In silico* identification and characterization of novel drug targets and outer membrane proteins in the fish pathogen *Edwardsiella tarda*. Open Access Bioinformatics 3, 37-42.

[17] Rathi, B., Sarangi, A.N., Trivedi, N. 2009. Genome subtraction for novel target definition in Salmonella typhi. Bioinformation 4, 143-150.

[18] Sakharkar, K.R., Sakharkar, M.K., Chow, V.T. 2004. A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa*. In Silico Biol 4, 355-360.

[19] Sarangi, A.N., Aggarwal, R., Qamar, R., Nidhi, T. 2009. Subtractive genomics approach for *in silico* identification and characterization of novel drug targets in Neisseria Meningitides Serogroup B. J Comput Sci Syst Biol 2, 255-258.

[20] Seltmann, G.H.O. 2002. The Bacterial Cell Wall. Springer Berlin, Heidelberg.

[21] Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., Jensen, L.J., von Mering, C. 2011. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. Nucl Acid Res 39, D561-D568.

[22] Udwadia, Z.F., Amale, A.R., Ajbani, K.K., Rodrigues, C. 2011. Totally drug-resistant tuberculosis in India. Clin Infect Dis, doi: 10.1093/cid/cir889.

[23] Victor, T.C., de Haas, P.E., Jordaan, A.M., van der Spuy, G.D., Richardson, M., van Soolingen, D., van Helden, P.D., Warren, R. 2004. Molecular characteristics and global spread of *Mycobacterium tuberculosis* with a western cape F11 genotype. J Clin Microbiol 42, 769-772.

[24] Warren, R., Richardson, M., van der Spuy, G., Victor, T., Sampson, S., Beyers, N., van Helden, P. 1999. DNA fingerprinting and molecular epidemiology of tuberculosis: Use and interpretation in an epidemic setting. Electrophoresis 20, 1807-1812.

[25] Warren, R.M., Streicher, E.M., Sampson, S.L., van der Spuy, G.D., Richardson, M., Nguyen, D., Behr, M.A., Victor, T.C., van Helden, P.D. 2002. Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: Implications for interpretation of spoligotyping data. J Clin Microbiol 40, 4457-4465.

[26] WHO (World Health Organization). 2010. Tuberculosis Fact sheet N°04.

[27] WHO (World Health Organization). 2011. Global Tuberculosis Control.

[28] Zhang, R., Ou, H.Y., Zhang, C.T. 2004. DEG: A database of essential genes. Nucl Acid Res 32, D271-D272.