International Journal of Mycobacteriology 2 (2013) 220–226

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/IJMYCO

Comparative proteomic analysis of *Mycobacterium tuberculosis* strain $H_{37}Rv$ versus $H_{37}Ra$

Lingaraja Jena [a], *Supriya Kashikar* [b], *Satish Kumar* [a], *Bhaskar C. Harinath* [a,*]

[a] *Bioinformatics Centre, Biochemistry & JBTDRC, MGIMS, Sevagram, Maharashtra, India*
[b] *GeNext Genomics Pvt. Ltd., Nagpur, Maharashtra, India*

ARTICLE INFO

Article history:
Received 7 September 2013
Received in revised form
11 October 2013
Accepted 14 October 2013
Available online 1 November 2013

Keywords:
MTB
Proteome
Virulence
Comparative proteomics
Bioinformatics

ABSTRACT

Background: *Mycobacterium tuberculosis* (MTB) $H_{37}Ra$ is an attenuated tubercle bacillus closely related to the virulent type strain MTB $H_{37}Rv$. In spite of extensive study, variation in virulence between the MTB $H_{37}Rv$ and MTB $H_{37}Ra$ strains is still to be understood. The difference in protein expression or structure due to mutation may probably be an important factor for the virulence property of MTB $H_{37}Rv$ strain.

*Methods:*In this study, a whole proteome comparison between these two strains was carried out using bioinformatics approaches to elucidate differences in their protein sequences.

*Results:*On comparison of whole proteome using NCBI standalone BLAST program between these two strains, 3759 identical proteins in both the strains out of 4003 proteins were revealed in MTB $H_{37}Rv$ and 4034 proteins were revealed in MTB $H_{37}Ra$; 244 proteins of MTB $H_{37}Rv$ and 260 proteins of MTB $H_{37}Ra$ were found to be non-identical. A total of 172 proteins were identified with mutations (Insertions/deletions/substitutions) in MTB $H_{37}Ra$ while 53 proteins of MTB $H_{37}Rv$ and 85 proteins of MTB $H_{37}Ra$ were found to be distinct. Among 244 non-identical proteins, 19 proteins were reported to have an important biological function; In this study, mutation was shown in these proteins of MTB $H_{37}Ra$.

*Conclusion:*This study reports the protein differences with mutations between MTB $H_{37}Rv$ and $H_{37}Ra$, which may help in better understanding the pathogenesis and virulence properties of MTB $H_{37}Rv$.

Introduction

Tuberculosis (TB) is a complex disease caused by *Mycobacterium tuberculosis* (MTB), which has evolved with highly successful mechanisms to equivocate host defenses and existing classes of antibiotics. Decades after the discovery of MTB, TB remains a major cause of morbidity and mortality in many developing countries. One third of the World's population is considered to be infected with MTB, with 8.7 million new patients and 1.4 million deaths in the year 2011, including 1 million deaths among HIV-negative and 430,000 HIV-positive individuals [1]. Multi-drug-resistant strains of this pathogen, emerging in association with HIV, have added a frightening dimension to the problem [2]. Outbreaks of extensively drug-resistant (XDR) TB have also been an increasing threat in certain regions around the world [3]. Despite abundant research on MTB diagnostics, vaccinations and treatments, this disease poses a considerable risk in many developed countries. MTB is very virulent, but there has been no simple answer found yet for what makes MTB so virulent. Historically, MTB $H_{37}Ra$ is the avirulent counterpart of the virulent strain MTB $H_{37}Rv$, and both strains were

* Corresponding author. Address: JB Tropical Disease Research Centre, Mahatma Gandhi Institute of Medical Sciences, Sevagram 442 102, Wardha, Maharashtra, India. Tel.: +91 7152 284341/355/262/303; fax: +91 7152 284038.
E-mail addresses: bc_harinath@yahoo.com, info@jbtdrc.org (B.C. Harinath).
2212-5531/$ - see front matter
http://dx.doi.org/10.1016/j.ijmyco.2013.10.004

derived from their virulent parent strain H37 discovered in 1935 by William Steenken through a process of aging and dissociation from *in vitro* culture [4]. These strains are phenotypically and genotypically different from each other, but the virulence power is different among these strains, which could probably be owing to a difference in protein expression. Owing to the advancement of Bioinformatics and the genome sequencing project, whole genome and proteome sequences of both the strains are available in the public domain. Current evidence suggests that as a species, MTB exhibits very little genomic sequence diversity [5,6]. MTB $H_{37}Rv$ is virulent and susceptible to most of the anti-tuberculous drugs, while MTB $H_{37}Ra$ is an avirulent strain and the MTB KZN (KwaZulu-Natal, South Africa) strain is resistant to different drugs like isoniazid, rifampicin, kanamycin, ofloxacin, ethambutol, pyrazinamide, etc. [7]. This may be due to a genetic mutation resulting in the generation of mutated proteins. Therefore, there is a need for genomic as well as proteomic analysis among different strains of MTB to understand the variation among them.

Many tools have also been developed for the complete determination of the genome sequence of a huge number of bacteria, but still, their proteomes remain relatively poorly defined. In the post-genomic era, proteomics is a rapidly growing field of research that is becoming increasingly important, because it deals with the study of proteins involved in carcinogenesis as well as a novel biomarker discovery for clinical use, such as screening, diagnosis, prognosis, detection of recurrent disease, etc. [8]. While a genome remains unchanged to a larger extent, the proteins in any particular cell change dramatically as genes are turned on or off in response to the environment.

Comparative genomic analysis of MTB $H_{37}Ra$ versus $H_{37}Rv$ by Zheng et al. revealed the genetic basis of virulence among these two strains [9]. However, proteomics is still nascent and requires extensive study. Since it is proteins that are directly involved in both normal and disease-related biochemical processes, a more comprehensive understanding of disease may be achieved by looking directly into the proteins within a disease cell or tissue [10]. Proteomics has much promise in novel drug discovery by targeting proteins of pathogenic organisms causing different diseases in the host, whereas comparative proteomics is very significant in studying the proteomic variations among different pathogens.

Identification of the virulence factors of MTB is a fundamental goal if new vaccines and anti-mycobacterial drugs against this pathogen are to be developed. A single amino acid mutation in protein sequence may cause alteration in the protein structure and function that may account for virulence and drug resistance properties of pathogenic organisms. Therefore, the development of an *in silico* technology to study the proteomic variations of different strains of genetically intractable pathogens such as MTB will enhance the analysis of virulence and drug resistance properties and significantly advance the understanding of the mechanisms of disease. In this study, the proteomic variations in these two strains (MTB $H_{37}Rv$ and $H_{37}Ra$) were determined, and the proteins that had undergone mutations (insertions/deletions/substitutions) were identified in the same variations. The findings of the present study provide a unique platform for the discovery of proteomic variation in other strains/species of *Mycobacterium* as well as the discovery and development of TB drugs, vaccines, biomarkers, etc.

Materials and methods

Dataset preparation

The dataset was prepared by retrieving the whole proteome of MTB $H_{37}Rv$ (NCBI RefSeq: NC_000962.2) and $H_{37}Ra$ (NCBI RefSeq: NC_009525.1) from the NCBI FTP site (ftp://ftp.ncbi.nih.gov/genomes/). The protein sequence data ($H_{37}Rv$ and $H_{37}Ra$ proteome) were formatted using in-house developed PERL Script for quick analysis. NCBI Standalone BLAST-2.2.26 (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/release/LATEST/) was used to perform protein BLAST [11] between the proteome of MTB $H_{37}Rv$ against MTB $H_{37}Ra$ and vice versa to discover protein variation and duplication.

Proteome comparison and database designing

Whole proteome comparison of these two organisms was done using PERL script and standalone BLAST. The output of the BLAST result was parsed and stored in MS SQL relational database tables using in-house developed PERL script. While parsing BLAST output results, percentage identities, positivities, number of gaps, identical residues, bits, bits score, e-value, query length, subject length, query sequence, subject sequence, consensus sequence, etc., of the first hit obtained were taken into consideration for each protein comparison.

Data retrieval and analysis

Different SQL queries were written to retrieve comparison data from MS SQL database tables. Sub-cellular localization (integral membrane, cytoplasmic, secretory and membrane attached by lipid anchor) of the selected MTB protein with variations were predicted by TBpred Prediction server [12].

Results

Genomic features and proteomic comparison of MTB $H_{37}Rv$ and its avirulent counterpart $H_{37}Ra$

MTB $H_{37}Rv$ contains a single circular chromosome of 4,411,532 bp with an average G + C content of 65.6% (NCBI RefSeq: NC_000962.2), which is 8445 bp smaller than the MTB $H_{37}Ra$ genome (NCBI RefSeq: NC_009525.1 and 4,419,977 bp length). A total of 4003 protein-coding sequences (CDS) are identified amongst 4062 genes in the $H_{37}Rv$ genome, while there are 4084 genes with 4034 protein-coding sequences in the genome of MTB $H_{37}Ra$ (Table 1). The proteomic comparison of MTB $H_{37}Rv$ and MTB $H_{37}Ra$ in this study revealed 3759 identical proteins between these two strains, while a reverse comparison, i.e., the proteome of MTB $H_{37}Ra$ against MTB $H_{37}Rv$, revealed 3774 identical proteins. Upon further analysis of this difference in number, it was found that there were 16 multiple copies of one protein, i.e., IS6110 transposase in MTB $H_{37}Ra$ while there were only four in MTB $H_{37}Rv$ of the same protein,

Table 1 – Genomic comparison between MTB $H_{37}Rv$ and MTB $H_{37}Ra$.

Feature	MTB $H_{37}Rv$	MTB $H_{37}Ra$
NCBI RefSeq ID	NC_000962.2	NC_009525.1
Genome length	4,411,532	4,419,977
Total genes	4062	4084
Total proteins	4003	4034

and a mutation was observed in the form of an amino acid deletion in the remaining 12 copies of IS6110 transposase.

There were also three extra proteins found to be duplicate in number in MTB $H_{37}Ra$ (MRA_1050∼MRA_1160, MRA_3421∼MRA_3514, MRA_2373∼MRA_2376), which were found as a single copy in MTB $H_{37}Rv$ (Rv1041c, Rv3474, Rv2352c). There were 244 proteins of MTB $H_{37}Rv$ found to be non-identical relative to MTB $H_{37}Ra$, and 260 proteins of MTB $H_{37}Ra$ were non-identical relative to MTB $H_{37}Rv$ (Table 2). Similar proteins with mutations (amino acid variations/substitutions, insertions and deletions of any size) and proteins without any significant similarity in their amino acid sequences among these two strains of MTB were considered non-identical proteins. Amongst 244 and 260 non-identical proteins, a total of 172 proteins were identified with mutations in MTB $H_{37}Ra$ when compared with MTB $H_{37}Rv$; 19 proteins with variations out of 244 in MTB $H_{37}Rv$ and three proteins out of 260 in MTB $H_{37}Ra$ may have no significant effect on the variation in the properties of these two strains as those proteins have extra copies in the respective strains. The rest of the proteins with respect to each strain were extensively varying in identities and query coverage. A few of them were unique to particular strains as they have no significant similarity against proteins of other strains like one hypothetical protein in MTB $H_{37}Ra$ (MRA_2157) and three hypothetical proteins (Rv0500B, Rv4012, Rv3599c) and two PE-PGRS family proteins (Rv3344c, Rv3512) in MTB $H_{37}Rv$.

Among those 172 mutated proteins, 46 were identified with amino acid substitutions, while 47 insertions, 64 deletions, one with insertion and deletion, four with insertions and substitutions, nine with deletions and substitutions and one with insertion, deletion and substitution in MTB $H_{37}Ra$ relative to MTB $H_{37}Rv$ (Table 2). Among the 46 proteins with amino acid variations, 40 have single amino acid mutations, while three with variations of two, two with four, and one with 12 amino acid substitutions were identified. There were 12 proteins of MTB $H_{37}Rv$ and 13 proteins of MTB $H_{37}Ra$ were found to have multiple copies.

Sub-cellular distribution of non-identical proteins with amino acid variation in MTB $H_{37}Ra$

A total of 172 proteins identified with variations in MTB $H_{37}Ra$ relative to MTB $H_{37}Rv$ were subjected to TBpred Prediction server [12] to discover their sub-cellular localization. It was found that most of the mutated proteins were either integral membrane proteins or cytoplasmic proteins. Out of the 172 proteins in MTB $H_{37}Rv$ 89 integral membrane proteins, 74 were cytoplasmic, six were secreted and three were attached to a membrane by a lipid anchor (Table 3).

Discussion

In spite of several studies in the past, the potential causes for variations in virulence between MTB $H_{37}Rv$ and MTB $H_{37}Ra$ have remained unclear. In this study, a comparative proteomic analysis was performed on two widely used MTB strains: $H_{37}Rv$ (virulent) and its avirulent counterpart $H_{37}Ra$, and their proteomic variations were studied through the Bioinformatics approach. As genome and proteome sequences of these two strains are available in the public domain, there is recently

Table 2 – Comparison of proteomic variations between MTB $H_{37}Rv$ and MTB $H_{37}Ra$.

Feature	MTB $H_{37}Rv$	MTB $H_{37}Ra$
Identical proteins in both the strains	3759	3774[a] (3759 + 12 + 3)
Non-identical proteins	244	260
Non-identical and distinct proteins	53	86
Non-identical and non-distinct proteins	19	2
Non-identical proteins with amino acid variation in MTB $H_{37}Ra$	172[b]	
a. Amino acid substitution	46	
b. Insertion (⩾1 AA)	47	
c. Deletion (⩾1 AA)	64	
d. Insertion/deletion	1	
e. Insertion/substitution	4	
f. Deletion/substitution	9	
g. Insertion/deletion/substitution	1	

[a] Proteomic comparison of MTB H37Rv and MTB H37Ra revealed 3759 identical proteins between these two strains while on reverse comparison, i.e., proteome of MTB H37Ra against MTB H37Rv revealed 3774 (3759 + 12 + 3) identical proteins. There were 16 multiple copies of one protein, i.e., IS6110 transposase in MTB H37Ra while there were only four in MTB H37Rv of the same protein and mutation was observed in the form of amino acid deletion in the remaining 12 copies of IS6110 transposase. There were also three extra proteins found to be duplicate in number in MTB H37Ra (MRA_1050∼MRA_1160, MRA_3421∼MRA_3514, MRA_2373∼MRA_2376) which were found as a single copy in MTB H37Rv (Rv1041c, Rv3474, Rv2352c).

[b] For mutation analysis MTB H37Rv was considered the reference strain. There were 172 proteins of MTB H37Ra which were identified with mutation relative to MTB H37Rv.

Table 3 – Sub-cellular distribution of non-identical proteins with amino acid variations.

Group	Total number
Integral membrane proteins	89
Cytoplasmic proteins	74
Secreted proteins	6
Proteins attached to membrane by lipid anchor	3

immense interest in analyzing the differences between the two strains in more detail.

Comparative genomic analysis of MTB $H_{37}Ra$ versus MTB $H_{37}Rv$ by Zheng et al. revealed 53 insertions and 21 deletions in MTB $H_{37}Ra$ relative to MTB $H_{37}Rv$ along with 198 single nucleotide variations (SNVs), 102 transitions, and 96 transversions in these two strains [9]. Malen et al. (2011) used label-free quantitative proteomic approach and estimated differences in protein abundance between these two strains [13]. They identified more than 1700 proteins from both strains, out of which 29 membrane-associated proteins were reported with five- or more-fold difference in their relative abundance in one strain compared with the other [13]. In a previous study, an MTB proteome comparison database (MTB-PCDB) was developed, which provides integrated access to protein sequence comparison with identical and non-identical protein data for five strains of MTB ($H_{37}Rv$, $H_{37}Ra$, CDC 1551, F11 and KZN 1435) [14]. The present study reports protein comparison of MTB $H_{37}Rv$ versus MTB $H_{37}Ra$ with respect to a detailed analysis of variation in different classes of proteins (secretory proteins, etc.) with mutation. This analysis identified 3759 identical proteins between these two strains, whereas 244 proteins of MTB $H_{37}Rv$ were found to be non-identical relative to MTB $H_{37}Ra$ along with some proteins with multiple copies in both the strains. Proteomic comparison information obtained from this study would be useful for a better understanding of the pathogenesis of MTB.

MTB comprises 11 serine/threonine protein kinases (STPKs) [15] which probably function in signal transduction pathways and may direct important cellular decisions such as dormancy and cell division [16,17]. Out of the 11 proteins, 10 proteins (Rv0014c, Rv0015c, Rv0410c, Rv0931c, Rv1743, Rv1746, Rv2088, Rv2176, Rv2914c and Rv3080c) were found to be identical in both the strains in this study, while a single amino acid substitution (R607Q) was identified in one of the proteins of MTB $H_{37}Rv$ (Rv1266c) compared with $H_{37}Ra$ (MRA_1274). Out of 19 proteins of MTB $H_{37}Rv$ reported as responsible proteins for cell division [16], only two, i.e., Rv0012 (probable conserved membrane protein) and Rv3919c (glucose inhibited division protein B), were found to have single amino acid substitutions, i.e., C233R and S100F, when compared with the proteins of MTB $H_{37}Ra$, i.e., MRA_0014 and MRA_3958, respectively, in this study.

Intensive research effort has shown that many proteins are associated with virulence factors of MTB [16], such as catalase-peroxidase [18], Mammalian cell entry (mce) operon [19], sigma factor gene (sigA) [20], Virulence Regulator EspR [21], Rv0485 regulon [22], Virulence Associated Protein (VapC) [23], PhoP regulon [24], etc. In this study, mutations in some of the virulent associated proteins were identified. Out of 33 mce family proteins of MTB $H_{37}Rv$, two proteins–Rv0175 and Rv0176–were found to have deletions in their amino acid sequences when compared with corresponding proteins of MTB $H_{37}Ra$, i.e., MRA_0183 and MRA_0184, respectively. Also, two proteins associated with virulence regulator EspR were identified with mutations, i.e., single amino acid substitution (G237V) in MRA_1078 related to Rv1068c (PE-PGRS family protein) and deletion of 57 amino acids in MRA_3527 related to Rv3487c (esterase). Besides, mutations were also observed in PhoP regulon and proteins associated with Rv0485 regulon (Table 4). Previous studies on non-identical proteins of MTB $H_{37}Rv$ by other researchers revealed the importance of those proteins (Table 4). This might account for the difference in the virulence and pathogenic properties among these two strains.

Mutations affecting integral membrane proteins

Most of the protein mutations were found in this category. A total of 89 proteins of MTB $H_{37}Ra$ identified with mutations relative to MTB $H_{37}Rv$ were predicted as integral membrane proteins. A deletion of 12 amino acids at N-terminal region was identified in the protein encoded by MRA_3163 in MTB $H_{37}Ra$ when compared with Rv3131 of MTB $H_{37}Rv$. This protein codes for a putative classical nitroreductase, which may be involved in detoxification of nitrogen-containing by-products in the host as reported by Purkayastha, et al. (2002) [25].

Rv3347c (PPE55) is reported to be recognized by antibodies elicited during subclinical TB infection of guinea pigs [26]. This protein when compared with proteins of $H_{37}Ra$ (MRA_3387) deletion of 12N-terminal amino acids was revealed.

Deletion of 3N-terminal amino acids in MRA_0763 with respect to Rv0754 was identified in the study. PE_PGRS11 protein encoded by Rv0754 recognizes TLR2 (toll-like receptor 2) and induces maturation and activation of Human Dendrite Cells [27]. Another PE-PGRS (Rv3367) protein has undergone mutation (insertion of three amino acids) in MRA_3407. This protein belongs to the PE protein family and has multiple tandem repeats of unique amino acid sequences and has characteristics of surface or secreted proteins like the other two proteins–PirG (Rv3810) and proline-threonine repetitive protein (PTRP) (Rv0538) [28].

It has been reported that Rv3654c and Rv3655c putative secreted proteins suppress apoptosis by blocking the extrinsic pathway [29]. These two proteins when compared with proteins of MTB $H_{37}Ra$ in this study also identified mutations (i.e., insertions in case of MRA_3689 and both insertion and deletion in case of MRA_3690).

Insertions of 80 amino acids in the C-terminal region of the nutrient stress–induced antigen Rv2660c (expressed during infection) [30] is identified when compared with MRA_2690 (hypothetical protein).

According to Goldstone et al., the *pe13* (Rv1195) and *ppe18* (Rv1196) gene pair is part of the Rv0485 regulon that is required for virulence of the pathogen during *in vivo* growth in mice [22]. In this study, it was also found that there were mutations (Insertions / Deletions / Substitutions) in the proteins of MTB $H_{37}Ra$, i.e., MRA_1370 and MRA_1205, when compared with the protein encoded by Rv1196 and Rv1195,

Table 4 – Mutation study of important functional proteins in MTB $H_{37}Ra$.

MTB $H_{37}Rv$	Importance/function	Ref.	MTB $H_{37}Ra$	Mutation type
Rv3131	May be involved in detoxification of nitrogen containing by-products in the host	[25]	MRA_3163	Deletion
Rv3347c	Recognized by antibodies elicited during subclinical TB infection of guinea pigs	[26]	MRA_3387	Deletion
Rv0754	Induce maturation and activation of human dendrite cells	[27]	MRA_0763	Deletion
Rv3367	Belongs to PE protein family with multiple tandem repeats of unique amino acid sequences	[28]	MRA_3407	Insertion
Rv3654c	Suppresses apoptosis by blocking the extrinsic pathway	[29]	MRA_3689	Insertion
Rv3655c	Suppresses apoptosis by blocking the extrinsic pathway	[29]	MRA_3690	Insertion/deletion
Rv2660c	Nutrient stress–induced antigen	[30]	MRA_2690	Insertion
Rv1195	Part of the Rv0485 regulon which is required for virulence of the pathogen	[22]	MRA_1205	Insertion
Rv1196	Part of the Rv0485 regulon which is required for virulence of the pathogen	[22]	MRA_1370	Deletion/insertion/substitution
Rv2829c	Virulence associated protein C (vapC), which might profoundly impact the physiology of the organism	[23]	No significant similarity found in MTB $H_{37}Ra$	
Rv0189c	Involved in valine, leucine and isoleucine biosynthesis	[31]	MRA_0197	Substitution
Rv1021	Required to maintain the full capacity of the mycobacterium to respond to oxidative stress	[32]	MRA_1029	Substitution
Rv0757	Involved in transcriptional mechanism. Part of the two component regulatory system PhoP/PhoQ.	[24]	MRA_0767	Substitution
Rv1099c	Encodes the major fructose-1,6-bisphosphatase II	[33]	MRA_1110	Deletion
Rv1248c	Essential for growth of MTB, involved in cellular metabolism. Has alpha-ketoglutarate dehydrogenase (KDH) activities.	[34]	MRA_1256	Deletion
Rv0496	T-cell antigen and potential vaccine candidate	[35]	MRA_0503	Deletion
Rv1738	Conserved hypothetical protein of unknown but essential function	[34]	MRA_1749	Insertion
Rv2466c	May play an important role in the SigH-regulated response to oxidative stress	[36]	No significant similarity found in MTB $H_{37}Ra$	
Rv0050	Encodes the bifunctional penicillin-binding protein ponA1 and essential for mycobacterial survival	[37]	MRA_0053	Substitution

respectively. Rv2829c is a hypothetical protein found to be unique in MTB $H_{37}Rv$ in the present study (as there is no significant similar protein in MTB $H_{37}Ra$). This protein was explored by Ahidjo et al. (2011) as a Virulence Associated Protein C (vapC) which belongs to the VapBC family that potentially provides an abundant source of RNase activity in MTB, which might profoundly impact the physiology of the organism [23].

Mutations affecting cytoplasmic proteins

A total of 74 proteins of MTB $H_{37}Ra$ identified with mutations were predicted as cytoplasmic proteins. Therefore, cytoplasmic proteins were found to be more prone to mutation, next to integral membrane proteins. Some of these proteins were found to be involved in important pathways of MTB.

It has been reported that the dihydroxy-acid dehydratase (DHAD) encoded by Rv0189c is essential for the survival of MTB and as it is absent in mammals, it could be a potential drug/vaccine target [31]. This protein is also known to be involved in five important pathways of MTB (Valine, leucine and isoleucine biosynthesis; Pantothenate and CoA biosynthesis; Biosynthesis of secondary metabolites; 2-Oxocarboxylic acid metabolism; Biosynthesis of amino acids) (http://www.genome.jp/dbget-bin/www_bget?mtu:Rv0189c). During this analysis, it was also found that this protein (Rv0189c) has undergone mutations of a single amino acid, i.e., Valine to Glycine at 284 position (V284G) in the corresponding protein of MTB $H_{37}Ra$ (MRA_0197).

Nucleoside triphosphate pyrophosphohydrolase enzyme encoded by Rv1021 is required to maintain the full capacity of the mycobacterium to respond to oxidative stress via the degradation of oxidation-induced damaged nucleotides [32]. It is known to be involved in two important pathways of MTB, i.e., Purine metabolism and Pyrimidine metabolism (http://www.genome.jp/dbget-bin/www_bget?mtu:Rv1021). The present analysis identified that a point mutation occurs at 219 position of the amino acid sequence of the corresponding protein of MTB $H_{37}Ra$ (MRA_1029), i.e., A219E, which is known to be important for catalytic activity. A mutation at this point reduced the activity of Pyrophosphohydrolase by 20-fold, which affects the magnesium binding and the protein structure [32].

The phoP encoded by Rv0757 is an essential protein with two-component response regulator activity. This analysis identified a point mutation S219L in the corresponding protein of MTB $H_{37}Ra$ (MRA_0767). Lee et al. (2008) also demonstrated that a mutation in the phoP of $H_{37}Ra$ results in an amino acid change from serine to leucine and is partially responsible for the decreased virulence of $H_{37}Ra$ [24].

Deletion of 22N-terminal amino acids was identified in the protein encoded MRA_1110 when compared with Rv1099c, which was reported to be expressed at significant levels in

MTB, and encodes the major fructose-1,6-bisphosphatase II (FBPase) of this pathogen [33]. This protein also found to be involved in six important pathways of MTB (http://www.genome.jp/dbget-bin/www_bget?mtu:Rv1099c).

Rv1248c is reported to be conserved in mycobacteria and other actinomycetes and predicted to be essential for growth of MTB [34]. It encodes multifunctional alpha-ketoglutarate metabolic enzymes and N-terminal deletion of 17 amino acids and is also identified in the corresponding protein (MRA_1256) of $H_{37}Ra$.

Rv0496 (MTB-PPX1) is reported as a T-cell antigen with potential for vaccine development [35]. This protein is known to be involved in purine metabolism pathways and has undergone N-terminal deletion of 16 amino acids along with amino acids variation in the avirulent counterpart $H_{37}Ra$ (MRA_0503). Rv1738 encodes a conserved hypothetical protein of unknown but essential function [34]. Insertion of 16 amino acids at the N-terminal end was identified in the corresponding proteins of $H_{37}Ra$ (MRA_1749).

There is a hypothetical protein (Rv2466c) found to be unique in MTB $H_{37}Rv$ in this study (as there is no significant similar protein in MTB $H_{37}Ra$). Raman et al. (2001) revealed the presence of glutaredoxin motif in Rv2466c and might play an important role in the SigH-regulated response to oxidative stress in MTB [36].

Mutations affecting secreted proteins and proteins attached to a membrane by lipid anchor

A total of 6 proteins of MTB $H_{37}Ra$ identified with mutations were predicted as secreted protein, while only three were predicted as proteins attached to a membrane by lipid anchor. The Rv0050 locus encodes the bi-functional penicillin-binding protein ponA1 and is essential to mycobacterial survival [37]. This analysis on this protein revealed a single amino acid mutation, i.e., Proline to Serine at 631 position (P631S) in the corresponding protein of MTB $H_{37}Ra$ (MRA_0197).

Conclusion

MTB $H_{37}Rv$ is virulent and susceptible to most of the antituberculous drugs, while MTB $H_{37}Ra$ is an avirulent strain. In spite of several studies in the past, the potential causes for variation in virulence between MTB $H_{37}Rv$ and MTB $H_{37}Ra$ have remained unclear. In this study, a comparative proteomic analysis of $H_{37}Ra$ against its virulent counterpart $H_{37}Rv$ was performed. These analyses provide proteomic differences between $H_{37}Rv$ and $H_{37}Ra$, which may useful for better understanding the basis of pathogenesis of MTB and virulence attenuation in MTB $H_{37}Ra$. Non-identical proteins identified in MTB $H_{37}Ra$ and MTB $H_{37}Rv$ must have some important role in the variation among these two strains directly or indirectly. 172 proteins were identified with mutations (Insertions/Deletions/Substitutions) and unique proteins identified in particular strains may be responsible for the variation. These proteins may be potential targets for the development of effective anti-mycobacterial strategy against this notorious pathogen and thus may be subject for further study, intending to carry out virulence determinants among the virulent and avirulent strains of MTB.

Acknowledgements

The authors thank the Department of Biotechnology, Ministry of Science & Technology, Government of India, for financial support to Bioinformatics Centre wherein this study has been carried out. The authors convey thanks to Shri Dhiru S Mehta, President, KHS, for his keen interest and encouragement.

REFERENCES

[1] World Health Organization, Global Tuberculosis Control, WHO Report 2012, 2012.
[2] M.C. Raviglione, D.E. Snider Jr., A. Kochi, Global epidemiology of tuberculosis. Morbidity and mortality of a worldwide epidemic, JAMA 273 (1995) 220–2266.
[3] N.S. Shah, A. Wright, G.H. Bai, L. Barrera, F. Boulahbal, N. Martín-Casabona, et al, Worldwide emergence of extensively drug-resistant tuberculosis, Emerg. Infect. Dis. 13 (2007) 380–387.
[4] W. Steenken Jr., L.U. Gardner, History of H37 strain of tubercle bacillus, Am. Rev. Tuberc. 54 (1946) 62–66.
[5] J.M. Musser, A. Amin, S. Ramaswamy, Negligible genetic diversity of *Mycobacterium tuberculosis* host immune system protein targets: evidence of limited selective pressure, Genetics 155 (2000) 7–16.
[6] S. Sreevatsan, X. Pan, K.E. Stockbauer, N.D. Connell, B.N. Kreiswirth, T.S. Whittam, et al, Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination, Proc. Natl. Acad. Sci. 94 (1997) 9869–9874.
[7] T.R. Ioerger, S. Koo, E.G. No, X. Chen, M.H. Larsen, W.R. Jacobs Jr, et al, Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa, PLoS One 4 (2009) e7778.
[8] W.C. Cho, Contribution of oncoproteomics to cancer biomarker discovery, Mol. Cancer 6 (2007) 25.
[9] H. Zheng, L. Lu, B. Wang, S. Pu, X. Zhang, G. Zhu, et al, Genetic basis of virulence attenuation revealed by comparative genomic analysis of *Mycobacterium tuberculosis* strain $H_{37}Ra$ versus $H_{37}Rv$, PLoS One 3 (2008) e2375.
[10] W.C. Cho, Proteomics technologies and challenges, Genomics Proteomics Bioinformatics 5 (2007) 77–85.
[11] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[12] M. Rashid, S. Saha, G.P. Raghava, Support vector machine-based method for predicting subcellular localization of mycobacterial proteins using evolutionary information and motifs, BMC Bioinformatics 8 (2007) 337.
[13] H. Målen, G.A. De Souza, S. Pathak, T. Søfteland, H.G. Wiker, Comparison of membrane proteins of *Mycobacterium tuberculosis* $H_{37}Rv$ and $H_{37}Ra$ strains, BMC Microbiol. 11 (2011) 18.
[14] L. Jena, G. Wankhade, S. Kumar, B.C. Harinath, MTB-PCDB: *Mycobacterium tuberculosis* proteome comparison database, Bioinformation 6 (2011) 131–133.
[15] M.J. Canova, R. Veyron-Churlet, I. Zanella-Cleon, M. Cohen-Gonsaud, A.J. Cozzone, M. Becchi, et al, The *Mycobacterium tuberculosis* serine/threonine kinase PknL phosphorylates Rv2175c: mass spectrometric profiling of the activation loop phosphorylation sites and their role in the recruitment of Rv2175c, Proteomics 8 (2008) 521–533.
[16] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, et al, Deciphering the biology of *Mycobacterium*

tuberculosis from the complete genome sequence, Nature 393 (1998) 537–544.
[17] Y. Av-Gay, J. Davies, Components of eukaryotic-like protein signaling pathways in *Mycobacterium tuberculosis*, Microb. Comp. Genomics 2 (1997) 63–73.
[18] S. Arruda, G. Bomfim, R. Knights, T. Huima-Byron, L.W. Riley, Cloning of an M. *tuberculosis* DNA fragment associated with entry and survival inside cells, Science 261 (1993) 1454–1457.
[19] A. Gioffré, E. Infante, D. Aguilar, M.P. Santangelo, L. Klepp, A. Amadio, et al, Mutation in mce operons attenuates *Mycobacterium tuberculosis* virulence, Microbes Infect. 7 (2005) 325–334.
[20] D.M. Collins, In search of tuberculosis virulence genes, Trends Microbiol. 4 (1996) 426–430.
[21] B. Blasco, J.M. Chen, R. Hartkoorn, C. Sala, S. Uplekar, J. Rougemont, et al, Virulence regulator EspR of *Mycobacterium tuberculosis* is a nucleoid-associated protein, PLoS Pathog. 8 (2012) e1002621.
[22] R.M. Goldstone, S.D. Goonesekera, B.R. Bloom, S.L. Sampson, The transcriptional regulator Rv0485 modulates the expression of a pe and ppe gene pair and is required for *Mycobacterium tuberculosis* virulence, Infect. Immun. 77 (2009) 4654–4667.
[23] B.A. Ahidjo, D. Kuhnert, J.L. McKenzie, E.E. Machowski, B.G. Gordhan, V. Arcus, et al, VapC toxins from *Mycobacterium tuberculosis* are ribonucleases that differentially inhibit growth and are neutralized by cognate VapB antitoxins, PLoS One 6 (2001) e21738.
[24] J.S. Lee, R. Krause, J. Schreiber, H.J. Mollenkopf, J. Kowall, R. Stein, et al, Mutation in the transcriptional regulator PhoP contributes to avirulence of *Mycobacterium tuberculosis* H_{37}Ra strain, Cell Host Microbe 3 (2008) 97–103.
[25] A. Purkayastha, L.A. McCue, K.A. McDonough, Identification of a *Mycobacterium tuberculosis* putative classical nitroreductase gene whose expression is coregulated with that of the acr gene within macrophages, in standing versus shaking cultures, and under low oxygen conditions, Infect. Immun. 70 (2002) 1518–1529.
[26] K.K. Singh, Y. Dong, S.A. Patibandla, D.N. McMurray, V.K. Arora, S. Laal, Immunogenicity of the *Mycobacterium tuberculosis* PPE55 (Rv3347c) protein during incipient and clinical tuberculosis, Infect. Immun. 73 (2005) 5004–5014.
[27] K. Bansal, S.R. Elluru, Y. Narayana, R. Chaturvedi, S.A. Patil, S.V. Kaveri, et al, PE_PGRS antigens of *Mycobacterium tuberculosis* induce maturation and activation of human dendritic cells, J. Immunol. 184 (2010) 3495–3504.
[28] K.K. Singh, X. Zhang, A.S. Patibandla, P. Chien Jr., S. Laal, Antigens of *Mycobacterium tuberculosis* expressed during preclinical tuberculosis: serological immunodominance of proteins with repetitive amino acid sequences, Infect. Immun. 69 (2001) 4185–4191.
[29] L. Danelishvili, Y. Yamazaki, J. Selker, L.E. Bermudez, Secreted *Mycobacterium tuberculosis* Rv3654c and Rv3655c proteins participate in the suppression of macrophage apoptosis, PLoS One 5 (2010) e10474.
[30] P.L. Lin, L. Dietrich, E. Tan, R.M. Abalos, J. Burgos, C. Bigbee, et al, The multistage vaccine H56 boosts the effects of BCG to protect cynomolgus macaques against active tuberculosis and reactivation of latent *Mycobacterium tuberculosis* infection, J. Clin. Invest. 122 (2012) 303–314.
[31] V. Singh, D. Chandra, B.S. Srivastava, R. Srivastava, Downregulation of Rv0189c, encoding a dihydroxyacid dehydratase, affects growth of *Mycobacterium tuberculosis in vitro* and in mice, Microbiology 157 (2011) 38–46.
[32] L.D. Lu, Q. Sun, X.Y. Fan, Y. Zhong, Y.F. Yao, G.P. Zhao, Mycobacterial MazG is a novel NTP pyrophosphohydrolase involved in oxidative stress response, J. Biol. Chem. 285 (2010) 28076–28085.
[33] F. Movahedzadeh, S.C. Rison, P.R. Wheeler, S.L. Kendall, T.J. Larson, N.G. Stoker, The *Mycobacterium tuberculosis* Rv1099c gene encodes a GlpX-like class II fructose 1,6-bisphosphatase, Microbiology 150 (2004) 3499–3505.
[34] C.M. Sassetti, D.H. Boyd, E.J. Rubin, Genes required for mycobacterial growth defined by high density mutagenesis, Mol. Microbiol. 48 (2003) 77–84.
[35] S. Bertholet, G.C. Ireton, M. Kahn, J. Guderian, R. Mohamath, N. Stride, et al, Identification of human T cell antigens for the development of vaccines against *Mycobacterium tuberculosis*, J. Immunol. 181 (2008) 7948–7957.
[36] S. Raman, T. Song, X. Puyang, S. Bardarov, W.R. Jacobs Jr., R.N. Husson, The alternative sigma factor SigH regulates major components of oxidative and heat stress responses in *Mycobacterium tuberculosis*, J. Bacteriol. 183 (2001) 6119–6125.
[37] L. Qin, J. Wang, R. Zheng, J. Lu, H. Yang, Z. Liu, et al, Perspective on sequence evolution of microsatellite locus (CCG)n in Rv0050 gene from *Mycobacterium tuberculosis*, BMC Evol. Biol. 11 (2011) 247.