ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES
Knowledge to Innovation

Vol 10, Issue 11, 2017

Online - 2455-3891
Print - 0974-2441

Research Article

EXPLOITATION OF HUB PROTEINS AS DRUG TARGETS FOR *MYCOBACTERIUM TUBERCULOSIS* H37RV

ASRA'A A ABDUL-JALIL*

University of Anbar Collage of Science .Email;suzimoh@yahoo.com

Received: 12 June 2017, Revised and Accepted: 09 August 2017

ABSTRACT

Objective: *Mycobacterium tuberculosis* (TB), a causative agent of TB, increased the resistance to most drugs in use and coinfection with HIV. It needs the new drug(s), the latter should be special candidate targets.

Methods: Hub proteins were studied (Rv2198c, Rv2507, and Rv3763) which had very high connected partners. These were modeled to be used for screening chemical databases.

Results: The proteins were found to be with increased low-complexity regions especially at the amino ends, they exhibited primary level (layers) of interactions and involved in secondary and more levels of interactions.

Conclusions: The proteins with high interacting partners would be good targets as their disruption will disturb the cellular functions. The chosen targets (Rv2198c, Rv2507, and Rv3763) have very high interactions at a different level (layers), they were modeled to be used in survey of chemical databases.

Keywords: *Mycobacterium tuberculosis*, Candidate drug targets, Interactome, Low-complexity regions.

 DOI: http://dx.doi.org/10.22159/ajpcr.2017.v10i11.20662

INTRODUCTION

Mycobacterium tuberculosis (TB) is an intracellular pathogen, a causative agent of TB, responsible for high rate of deaths, MTB faces a highly hostile environments during infection, such as limited nutrients, reduced oxygen tension [1], and immune system. MTB genome comprises 4,411,707 bp with about 4000 genes [2]. Having high guanine + cytosine content that reflected in biased amino acids in the proteins [2]. It expressed unique mechanisms to stand different stresses [3], through the proteins which are the main catalysts, structural components and signaling messengers, they are the machines of a biological system [4].

The situation with TB is worsened recently with the emergence of MDR, XDR strains and coinfection with HIV [1,5]. There is a need for therapeutics based on innovative drug targets [6,7], as about 70% of the failure across all therapeutics were attributed to the lack of efficiency [8]. So hub proteins represented good targets to be attacked, especially the highly connected hubs as it disturbance leads to a rapid disruption of communication through the cellular network [9,10]. This depends on the fact that a drug gable target is a protein, peptide or nucleic acid with an activity that can be modulated by a drug or chemicals [11]. High protein connections could be attributed to the presence of low-complexity regions (LCRs), which believed to play pivotal roles across a wide range of biological functions [12,13]. It has been noticed that the highly interacting proteins contain an increased fraction of LCRs compared to non-hub proteins [14].

It is known that there are several strategies exist to pursuit of drug design [8], in this study hub proteins were studied to be new drug targets for MTB.

METHODS

The protein sequence of Rv2198c (mmpS3), Rv2507 (hypothetical protein), and Rv3763(LpqH) of MTB H37Rv used in this study, the proteins were prepared to be used as drug targets.

Protein characterization

1. Amino acids sequences were retrieved from different TB database(s).
2. Essentiality was checked using DEG database (http://www.essentialgene.org/ [15], and TDR target database V 5.0 (http://tdrtargets.org/ [16].
3. Secondary structures were estimated using GOR4 server https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html [17].
4. Interactome was estimated using STRING database v 10 http://string-db.org/ [18], for 1st layer of interactions, visANT tool was used for the 2nd layer of interactions. http://visant.bu.edu. [19].
5. Amino acids similarity of proteins was estimated using BLASTp program https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome [20,21].
6. Identification of sequences was performed using http://www.essentialgene.org/ [15], and TDR target database V 5.0 http://tdrtargets.org/.
7. Cellular location of proteins estimated using Hidden Markov Model (TMHMM) http://www.cbs.dtu.dk/services/TMHMM/ [22].
8. Functionality estimated using pdbsum database http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html [23].
9. LCR were identified using SEG software incorporated in SMART database http://smart.embl-heidelberg.de/ [24].
10. Other protein characterization mostly determined according to TubercuList database http://tuberculist.epfl.ch/quicksearch.php?gene+name=Rv2198c&submit=Search [2].

RESULTS AND DISCUSSION

In this study, three proteins were chosen as they have a high degree of interactions with other proteins either directly or indirectly. These were Rv2198c which is connected to about 1573 proteins; Rv2507 connects to 1963 proteins, and Rv3763 connects to 2306 proteins (http://www.imtech.res.in/raghava/mycoprint/), to be used as drug targets, the following shows Protein Characterization.

Rv2198c

Product protein mmpS3, consist of 299 amino acids (900 bp) [2]. BLASTing of protein sequence revealed that this protein is exclusive for *Mycobacterium* genus on screening 1000 genomes using E value 1e-20 and word size 6, this confirmed by searching TARGET website (http://webhost.nts.jhu.edu/target/search.aspx?query [25]).

Cellular location of the protein shown in the following Fig. 1.

The secondary structure estimation showed that the protein composed of random coil structure (73.58%) and β-sheet to some extent (results not shown).

The protein is non-essential according to DEG database [15], and this is not true for *in vivo* essentiality, since the essentiality in DEG build on the results of *in vitro* experiments, i.e., growth on/in lab media, this confirmed by results of TDR target database [16], which also indicates that this protein belongs to functional category "cell wall and cell process" according to Gene Ontology, and the gene in group of low level expression (0-20%), and expressed in slow growth in the dormant phase of bacterial life cycle and is essential under these situation, in addition with low drug gable level (0.3).

Rv2507

A hypothetical protein (MT2582), consists of 273 amino acids (822 bp), BLASTing of sequence and revision of TARGET website [16] indicated that the protein is mainly mycobacterial, and shows low similarity (27-29%) with some species of *Gordonia* genera and to less extent to *Nocardia*, when screening 1000 genomes included in NCBI. Cellular location revealed that it is membrane associated as shown in Fig. 2.

The secondary structure investigation showed that is mainly 68.86% is random coils (results not shown). The protein is essential for *in vitro* growth [15,16]. Database(s) and gene ontology approach indicate that it belongs to functional category "cell wall and cell process," it is grouped with highly expressed genes (80-100%), and expressed in dormant phase, with low druggability level (0.3).

Rv3763

A protein (lpqH, MTV025.11), composed of 159 amino acids (480 bp), BLASTing and survey of 1000 genomes of NCBI databases showed that is mainly mycobacterial, and showed some similarity extended to about 40-50% for *Gordonia* genus, which is confirmed by results of TARGET website [16], cellular location associated with membrane as shown in Fig. 3.

Protein with PDB structure (4zjm), composed of β-sheets (45.91%) and random coils at (54.09) (results not shown). The gene (protein) is non-essential for *in vitro* growth of MTB H37Rv [15]. Gene Ontology and pdbsum database results revealed that the protein belongs to functional category "cell wall and cell process," the gene expressed at low to moderate level (20-40%), it is upregulated in a dormant phase, no chemicals were assigned to react with this protein [16].

Interactions and LCR

The analysis of pathogen interactome is a powerful approach for dissecting potential pathways and offers opportunities to be explored as new drug targets [13]. PPIs of the studied proteins were studied, the first layer of interactions estimated using STRING database as shown in the following Fig. 4.

The second and third layers of interaction estimated using visANT server as shown in the following Fig. 5.

PPIs forms the basis of cellular events such as signaling, regulation, and other processes [3,26], therefore the most highly connected proteins are usually the most important and are considered to participate extensively in cellular processes [27] these comprises networks that can be predicted computationally that leverage the large amount of

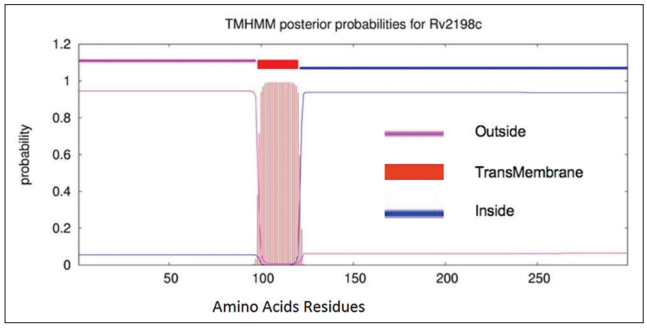

Fig. 1: Membrane association of Rv2198c protein

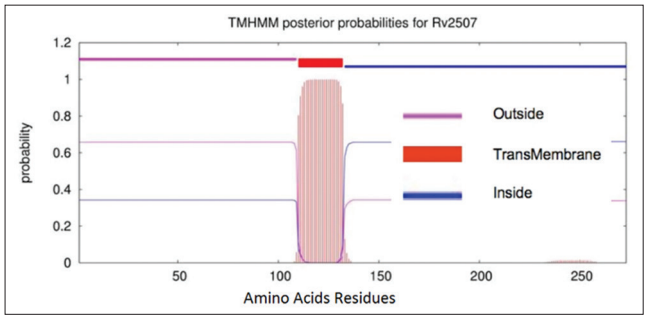

Fig. 2: Membrane association of Rv2507 protein

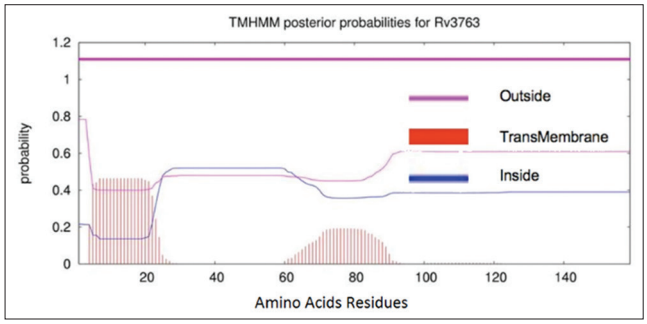

Fig. 3: Membrane association of Rv3763 protein

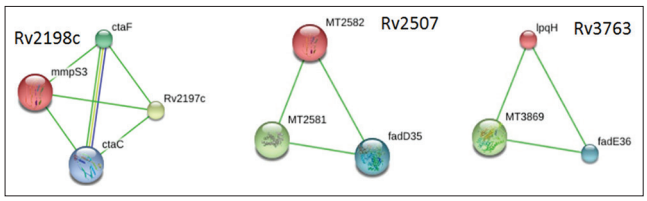

Fig. 4: Primary interactions of Rv2198c, Rv2507, and Rv3763

sequences data generated recently [26]. It would be expected that the highly connected proteins are chaperones, especially in MTB facing hostile environmental factors [3]. Due to the importance of PPIs, many databases were built up, and computational based on protein sequence and structures have been developed and widely used [26].

Fortunately, most databases and software do combine multiple methods (as in visANT software) for predicting interactions using different methods, visANT software depends on Predictome database and adapts different methods for PPIs prediction [19,26]. It is known that selective targets governs the drug discovery process, and it is quite useful in analyzing how node deletions in a network can disrupt the flow of information which helps in drug discovery and drug resistance pathways [11,28], this indicates that despite those established drug

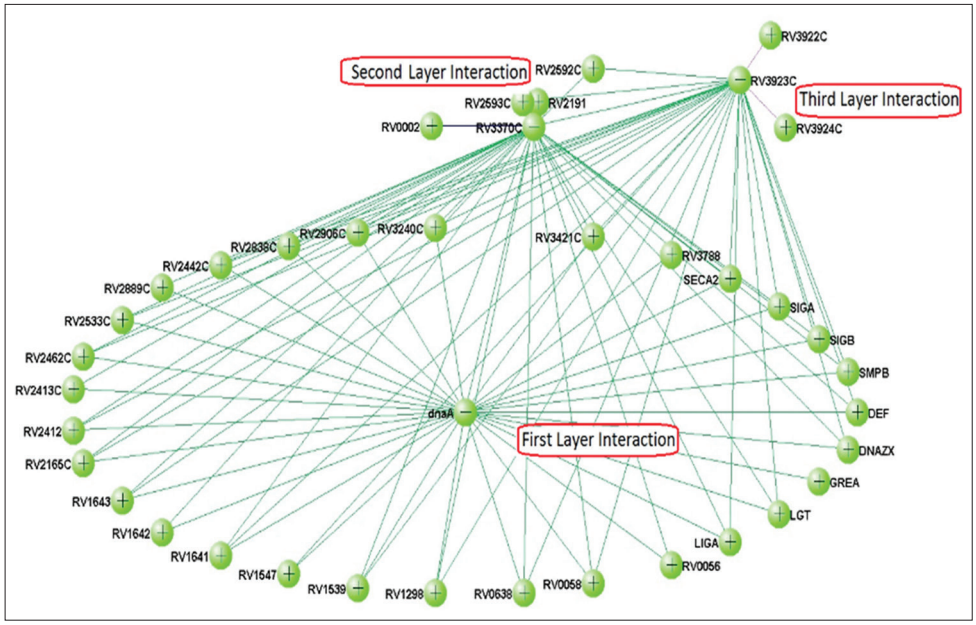

Fig. 5: Different layers of interaction of *Mycobacterial* proteins

Fig. 6: (a-c) Low-complexity regions in the studied protein sequences

target classes, innovative approaches are addressing previously undruggable target classes such as PPIs could be exploited [11].

However, PPIs network building software and their infrastructures databases are some flaws, therefore complete network sometimes cannot be perfectly obtained, so building our network divided into two layers as shown above.

In general, the hub proteins contain LCRs (amino acids sequences that contain repeats of single amino acid or short amino acids motif), these regions enable them to have more binding patterns across different PPIs networks than proteins that have no LCRs [12], the following figures show the LCRs of the studied proteins.

LCRs may be involved in flexible binding associated with specific functions depending on their position in protein sequence which is important for binding properties and biological roles [11,13], it was found that some bacteria such as MTB, *Pseudomonas aeruginosa*, *Deinococcus radiodurans,* and others have high portion of LC sequences in their genomes [13,29], which led to suggestion that LCRs are common sources of genetic variation in prokaryotes and can be contribute to the formation of novel coding sequences facilitating the generation of novel

protein functions to adapt to fast evolving environments [13,30], this evidence comes from the finding that recently emerged proteins contain more LC sequences than older proteins and these sequences often form functional domains [31]. The presence of LCRs which is flexible regions lacking well-defined folding structures is thought to be responsible for their versatile binding capabilities and allow them to bind several different targets [12,32] as appeared with the studied proteins which mostly having no very well organized secondary structures. In addition, it has been noticed that the highly connected hub proteins contain an increased fraction of LCRs compared to non-hub proteins [14] as shown in the studied proteins (Fig. 6). In addition, the increased ability of the studied hub proteins (this study) might be explained by the fact that the proteins with LCRs in their sequence extremities (t-LCRs) have more binding partners than proteins with central LCRs (c- LCRs), it has been found that the length of LCRs are positively correlated with the number of binding partners especially those in the sequence extremities [12].

Finally, it is obvious that to combat TB, both active replicating bacteria and dormant non- replicating should be eliminated [7,33], and this compatible with new approaches in nanomedicine and the progress in the comparative docking studies [34,35]. Anyway, the studied protein was found to be expressed in dormant phase of the bacterial growth cycle, being thought that it is a good strategy to attack the TB. 3D structure was estimated, and structure - based inhibitors screening, and prediction of binding sites for chemicals or drugs are going on.

CONCLUSION

The new drug targets of MTB need to be very effective to disrupt the cellular functions. The proteins with high interacting partners would be good targets as their disruption will disturb the cellular functions. The chosen targets (Rv2198c, Rv2507, and Rv3763) have very high interactions at a different level (layers), they were modeled to be used in a survey of chemical databases.

REFERENCES

1. James PE, Grinberg OY, Michaels G, Swartz HM. Intraphagosomal oxygen in stimulated macrophages. J Cell Physiol 1995;163(2):241-7.
2. Tekaia F, Gordon SV, Garnier T, Brosch R, Barrell BG, Cole ST. Analysis of the proteome of *Mycobacterium tuberculosis in silico*. Tuber Lung Dis 1999;79(6):329-42.
3. Cui T, Zhang L, Wang X, He Z. Uncovering new signaling proteins and potential drug targets through the interactome analysis of *Mycobacterium tuberculosis*. BMC Genomics 2009;10:118-27.
4. Eisenberg D, Marcotte EM, Xenarios I, Yeates TO. Protein function in the post-genomic era. Nature 2000;405(6788):823-6.
5. Kanji A, Hasan R, Zaver A, Ali A, Imtiaz K, Ashraf M, *et al*. Alternate efflux pump mechanism may contribute to drug resistance in extensively drug-resistant isolates of *Mycobacterium tuberculosis*. Int J Mycobacteriol 2016;5 Suppl 1:S97-8.
6. Koseki Y, Kanetaka H, Tsunosaki J, Munier-Lehmann H, Aoki S. Tetrahydro-2-furanyl-2,4(1H,3H)-pyrimidinedione derivatives as novel antibacterial compounds against *Mycobacterium*. Int J Mycobacteriol 2017;6:61-9.
7. Korycka-Machala M, Nowosielski M, Kuron A, Rykowski S, Olejniczak A, Hoffmann M, *et al*. Naphthalimides selectively inhibit the activity of bacterial, replicative DNA ligases and display bactericidal effects against tubercle bacilli. Molecules 2017;22:1-10.
8. Crowther GJ, Shanmugam D, Carmona SJ, Doyle MA, Hertz-Fowler C, Berriman M, *et al*. Identification of attractive drug targets in neglected-disease pathogens using an *in silico* approach. PLoS Negl Trop Dis 2010;4(8):e804.
9. Newman M. The structure and function of complex networks. SIAM Rev 2003;45:167-256.
10. Albert R, Jeong H, Barabási A. Error and attack tolerance of complex networks. Nature 2000;406:378-82.
11. Hartkoorn RC, Sala C, Neres J, Pojer F, Magnet S, Mukherjee R, *et al*. Towards a new tuberculosis drug: Pyridomycin-nature's isoniazid. EMBO Mol Med 2012;4(10):1032-42.
12. Coletta A, Pinney JW, Solís DY, Marsh J, Pettifer SR, Attwood TK. Low-complexity regions within protein sequences have position-dependent roles. BMC Syst Biol 2010;4:43.
13. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic tandem repeats generate functional variability. Nat Genet 2005;37(9):986-90.
14. Ekman S, Bjorklund A, Elofsson A. What properties characterize the hub proteins of the protein-protein interaction network of the protein interaction network of *Saccharomyces cerevisiae*? Genome Biol 2006;7:R45.
15. Zhang R, Lin Y. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. Nucleic Acids Res 2009;37:D455-8.
16. Magariños MP, Carmona SJ, Crowther GJ, Ralph SA, Roos DS, Shanmugam D, *et al*. TDR Targets: A chemogenomics resource for neglected diseases. Nucleic Acids Res 2012;40:D1118-27.
17. Garnier J, Gibrat JF, Robson B. GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol 1996;266:540-53.
18. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, *et al*. STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015;43:D447-52.
19. Hu Z, Mellor J, Wu J, DeLisi C. VisANT: An online visualization and analysis tool for biological interaction data. BMC Bioinformatics 2004;5:17.
20. Altschul S, Madden T, Schäffer A, Zhang J, Zhang Z, Miller W, *et al*. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 1997;25:3389-402.
21. Altschul SF, Wootton J, Gertz M, Agarwala R, Morgulis A, Schäffer A, *et al*. Protein database searches using compositionally adjusted substitution matrices. FEBS J 2005;272:5101-9.
22. Krogh A, Larsson B, von Heijne G, Sonnhammer E. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol 2001;305:567-80.
23. Laskowski RA. PDBsum: Summaries and analyses of PDB structures. Nucleic Acids Res 2001;29:221-2.
24. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: A web-based tool for the study of genetically mobile domains. Nucleic Acids Res 2000;28:231-4.
25. Lamichhane G, Zignol N, Blades D, Geiman A, Dougherty J, Grosset K, *et al*. A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: Application to *Mycobacterium tuberculosis*. Proc Natl Acad Sci USA 2003;100:7213-8.
26. Raman K. Construction and analysis of protein-protein interaction networks. Autom Exp 2010;2(1):2.
27. Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. Nature 2001;411(6833):41-2.
28. Raman K, Chandra N. *Mycobacterium tuberculosis* interactome analysis unravels potential pathways to drug resistance. BMC Microbiol 2008;8:234.
29. Wanker EE, Sun Y, Savitz AJ, Meyer DI. Functional characterization of the 180-kD ribosome receptor *in vivo*. J Cell Biol 1995;130(1):29-39.
30. Moxon ER, Rainey PB, Nowak MA, Lenski RE. Adaptive evolution of highly mutable loci in pathogenic bacteria. Curr Biol 1994;4(1):24-33.
31. Toll-Riera M, Radó-Trilla N, Martys F, Albà MM. Role of low-complexity sequences in the formation of novel protein coding sequences. Mol Biol Evol 2012;29(3):883-6.
32. Dyson H, Wright P. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 2005;6:197-208.
33. Iacobino A, Piccaro G, Giannoni F, Mustazzolu A, Fattorini L. Activity of drugs against dormant *Mycobacterium tuberculosis*. Int J Mycobacteriol 2016;5 Suppl 1:S94-5.
34. Shaji J, Shaikh M. Drug-resistant tuberculosis: Recent approach in polymer based nanomedicine. Int J Pharm Pharm Sci 2016;8(10):1-6.
35. Kaithamanakallam R, Karunakaran R, Srikumar P. Anti-tuberculosis drugs against mycobacterium tuberculosis PknB and mutants L33d/D76A-a comparative docking study. Int J Pharm Pharm Sci 2014;6(1):662-4.