Physico-chemical characterization (in silico analysis) of Rv2034 protein, a transcription factor from Mycobacterium tuberculosis

Bhanu Pratap
Research Scholar
School of Life Sciences,
Jawaharlal Nehru University, New Delhi-67 (India)

Dr.Arvind Lal
Assistant Professor
Department of Chemistry,
Central University of Jharkhand,Brambe, Ranchi - 835205,(India)

Abstract- According to the world health organization statistics, approx. 10.4 million people are infected with tuberculosis each year and 1.8 million people died from the disease. It is observed that the latent stage of mycobacteria has infected more than one-third of the world's population. The latent mycobacteria (MTB) is a major health concern, as mycobacteria remain dormant for long period and require longer treatment to eradicate the disease. A large no of gene regulatory proteins are synthesized by *Mycobacterium tuberculosis*. Gene regulation has an important role during MTB infection because it helps in persistence and virulence. Rv2034, a transcription factor involved in the upregulation of PhoP. In the present study, we have analyzed the physio-chemical properties of Rv2034 protein by using the Protparam tool and found that it is a hydrophobic and stable protein. The Secondary structure analysis was done using PredictProtein, RaptorX, HNN, PHD, and SOPMA using, online server and it was found that protein is rich in α-helical content.

Keywords: *Mycobacterium tuberculosis*, Transcriptional regulator, Primary structure, Virulence, Pathogenesis

I. INTRODUCTION

Tuberculosis (TB) has become one of the top ten causes of death worldwide. Around 30% of the population is infected with *Mycobacterium tuberculosis* (MTB) and causes millions of death per year[1]. MTB is an extremely successful pathogen because of its ability to persist in infected individuals without producing any symptom. Studies on Mtb have indicated that PhoP play important role in virulence and pathogenesis[2]. It regulates the variety of genes involved in hypoxia responses, respiratory metabolism, secretion of the major T-cell antigen ESAT-6, stress responses, synthesis of pathogenic lipids, and M. tuberculosis persistence[3, 4]. Rv2034 is a member of ArsR- type transcriptional regulators[5]. It acts as an auto repressor and, positively regulates the expression of *dosR*, *phoP*, and *groEL2*[6]. Rv2034 binds to a palindrome sequence motif present on target genes[7]. It is involved in the regulation of Metal stress by modulating the expression of genes related to metal uptake, efflux, sequestration, or detoxification[8, 9]. In the present study, Rv2034 is analyzed by bioinformatics tools. Based on the analysis, Rv2034 is a hydrophobic and stable protein. Secondary structure analysis shows that protein is rich in α-helical content.

II. METHODOLOGY

2.1. Sequence retrieval and primary structure.

The amino acid sequence (Met1-Thr107) of Rv2034 from Mycobacterium tuberculosis H37Rv (UniProt id: O53478) was retrieved from the UniProtKB database to analyze their physiochemical, structural properties. Gene sequence and amino acid sequences were also retrieved from NCBI (National Center for Biotechnology Information) database (http://www.ncbi.nlm.nih.gov) in FASTA format for computational analysis. The conserved domain identification was done using the conserved domain database (CDD)[10]. The primary structure and the amino acid

composition of Rv2034 were computed using Protparam tools[11]. Amino acid sequence retrieved from UniProtKB database was used as the input sequence.

2.2. Secondary structure prediction.

The secondary structure of Rv2034 was predicted by using online tools like PredictProtein[12], RaptorX[13], HNN[14], PHD[15], and SOPMA[16] and the secondary amount was calculated.

2.3. Physio-chemical properties.

The amino acids of Rv2034 protein sequences contain various information such as isoelectric point (pI), molecular weight (Mw), extinction coefficient (Ec), instability index (II), aliphatic index (AI), and Grand average of hydropathicity (GRAVY). All the physiochemical properties were calculated from the ProtParm tool.

III. RESULTS AND DISCUSSION

3.1. Primary structure

Rv2034 gene is located upstream of Rv2033 and downstream of the Rv2035 gene in the Mycobacterium tuberculosis genome. The conserved domain database (CDD) analysis shows that this contains HTH DNA binding domain (Figure 1). The gene sequence is as follows.

>NC_000962.3:2281294-2281617 *Mycobacterium tuberculosis* H37Rv
GTGTCCACTTACAGATCACCGGATCGCGCTTGGCAGGCGCTGGCGGACGGCACTCGCCGGGCCATCGTGG
AGCGGCTGGCGCACGGCCCGCTGGCCGTCGGCGAGTTGGCCCGCGACCTGCCCGTCAGCCGACCCGCGGT
GTCACAGCACCTCAAAGTGCTCAAGACCGCCAGGCTGGTGTGCGACCGCCCCGCGGGAACACGCCGCGTC
TACCAGCTCGACCCGACAGGCCTTGCGGCATTGCGCACCGACCTCGACCGGTTCTGGACACGCGCCCTGA
CTGGCTACGCGCAGCTCATCGACTCCGAAGGAGACGACACATGA

The Rv2034 is a 107 amino acid long protein(Met1toThr107). The amino acid sequence was retrieved from UniprotKB which is as follows. Details of the composition of amino acids are mentioned in Table 1. Alanine, arginine, and leucine are predominantly present in the protein. Details of amino acid composition are as follows.

>sp|O53478|HTHAR_MYCTU HTH-type transcriptional regulator Rv2034
MSTYRSPDRAWQALADGTRRAIVERLAHGPLAVGELARDLPVSRPAVSQHLKVLKTARLVCDRPAGTRRVYQLDPTG
LAALRTDLDRFWTRALTGYAQLIDSEGDDT

Table 1- Composition of amino acids

Name of Amino acid	No. of molecule present	Percentage
Ala (A)	14	13.1%
Arg (R)	14	13.1%
Asn (N)	0	0.0%
Asp (D)	10	9.3%
Cys (C)	1	0.9%
Gln (Q)	4	3.7%
Glu (E)	3	2.8%
Gly (G)	7	6.5%
His (H)	2	1.9%
Ile (I)	2	1.9%
Leu (L)	14	13.1%
Lys (K)	2	1.9%
Met (M)	1	0.9%
Phe (F)	1	0.9%
Pro (P)	6	5.6%
Ser (S)	5	4.7%
Thr (T)	9	8.4%
Trp (W)	2	1.9%
Tyr (Y)	3	2.8%
Val (V)	7	6.5%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

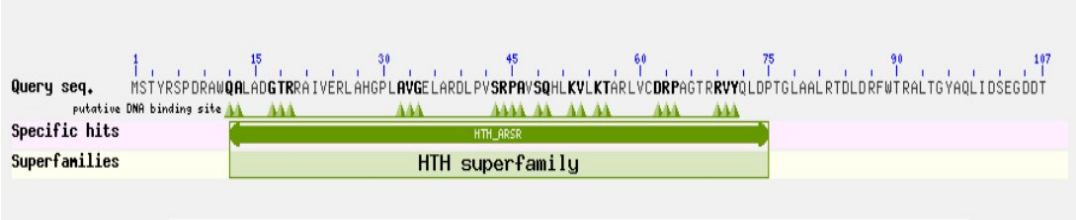

Figure1. Shows HTH DNA binding domain analyzed by using the conserved domain database (CDD)

3.2. Physiochemical properties :

Physio-chemical characterization is very important to characterize specific proteins. The average Molecular weight of Rv2034 proteins calculated is 11856.47 dalton. ProtParam tool was used to study the physiochemical properties of Rv2034 proteins. The results are shown in Table 2. Protein pI is calculated using the pKa values of amino acids.

Table 2. Physio-chemical Properties of Rv2034

Properties	Value
No. of Amino acids	107
Mol. Wt. in Dalton	11856.47
Theoretical pI	9.50
Asp+Glu	13
Arg+Lys	16
Extinction coefficients (EC)	15470
Aliphatic index	90.37
GRAVY	-0.384

3.3. *Secondary structure prediction*

The secondary structure of Rv2034 protein was predicted using online tools like PredictProtein, RaptorX, HNN, PHD, and SOPMA. The percentage of α-Helix, β-Sheets, and Random coil was calculated and it was found that Rv2034 protein is rich in α- helical content. The secondary structure content of proteins was predicted by online tools are mentioned in Table 3.

Table-3 Secondary structure content of Rv2034

Tools	α-Helix (%)	β-Sheets (%)	Random coil (%)
PredictProtein	63.55	8.41	28.04
RaptorX	56	10	33
HNN	56.07	9.35	34.58
PHD	59.81	10.28	29.91
SOPMA	55.14	7.48	28.97

III. CONCLUSION

Rv2034 is a 11856.47 Dalton protein and forms homodimer in solution. Primary structure analysis shows that Rv2034 is hydrophobic due to the high content of arginine and lysine residues (Arg+Lys=16). The aliphatic index (90.37)computed by Expasy's ProtParm infers that this protein is stable for a wide range of temperatures. Secondary structure analysis shows that Rv2034 proteins have predominant α-helical structures.

REFERENCES

[1] A.M. Ginsberg, M. Spigelman, " Challenges in tuberculosis drug research and development ", Nature medicine 13(3), pp 290-294., 2007.

[2] S.B. Walters, E. Dubnau, I. Kolesnikova, F. Laval, M. Daffe, I. Smith, "The *Mycobacterium* tuberculosis PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis" , Molecular microbiology 60(2) pp 312-330, 2006.

[3] W. Frigui, D. Bottai, L. Majlessi, M. Monot, E. Josselin, P. Brodin, T. Garnier, B. Gicquel, C. Martin, C. Leclerc, " Control of M. tuberculosis ESAT-6 secretion and specific T cell recognition by PhoP, PLoS Pathog 4(2) e33 ,2008.

[4] J. Gonzalo-Asensio, S. Mostowy, J. Harders-Westerveen, K. Huygen, R. Hernández-Pando, J. Thole, M. Behr, B. Gicquel, C. Martín, " PhoP: a missing piece in the intricate puzzle of *Mycobacterium tuberculosis* virulence", PloS one 3(10)) e3496. 2008

[5] C. Martin, A. Williams, R. Hernandez-Pando, P.J. Cardona, E. Gormley, Y. Bordat, C.Y. Soto, S.O. Clark, G.J. Hatch, D. Aguilar, " The live *Mycobacterium tuberculosis* phoP mutant strain is more attenuated than BCG and confers protective immunity against tuberculosis in mice and guinea pigs" , Vaccine 24(17) pp ,3408-3419, 2006.

[6] C.-H. Gao, M. Yang, Z.-G. He, " An ArsR-like transcriptional factor recognizes a conserved sequence motif and positively regulates the expression of phoP in mycobacteria" , Biochemical and biophysical research communications 411(4) pp 726-731, 2011.

[7] C.-h. Gao, M. Yang, Z.-G. He, " Characterization of a novel ArsR-like regulator encoded by Rv2034 in *Mycobacterium tuberculosis"* , PloS one 7(4)) e36255,2012.

[8] A.O. Summers, " Damage control: regulating defenses against toxic metals and metalloids" , Current opinion in microbiology 12 (2),pp 138-144, 2009.

[9] S. Tottey, D.R. Harvie, N.J. Robinson, " Understanding how cells allocate metals using metal sensors and metallochaperones" , Accounts of chemical research 38(10), pp775-783, 2005

[10] A. Marchler-Bauer, S. Lu, J.B. Anderson, F. Chitsaz, M.K. Derbyshire, C. DeWeese-Scott, J.H. Fong, L.Y. Geer, R.C. Geer, N.R. Gonzales, " CDD: a Conserved Domain Database for the functional annotation of proteins" , Nucleic acids research 39(suppl_1), pp225-229, 2010.

[11] E. Gasteiger, C. Hoogland, A. Gattiker, M.R. Wilkins, R.D. Appel, A. " Bairoch, Protein identification and analysis tools on the ExPASy server " , The proteomics protocols handbook" , Springer, pp. 571-607, 2005

[12] G. Yachdav, E. Kloppmann, L. Kajan, M. Hecht, T. Goldberg, T. Hamp, P. Hönigschmid, A. Schafferhans, M. Roos, M. Bernhofer, " PredictProtein—an open resource for online prediction of protein structural and functional features" , Nucleic acids research 42(W1) W337-W343, 2014.

[13] S. Wang, J. Peng, J. Ma, J. Xu, " Protein secondary structure prediction using deep convolutional neural fields" , Scientific reports 6(1) pp1-11, 2016.

[14] K. Lin, V.A. Simossis, W.R. Taylor, J. Heringa, " A simple and fast secondary structure prediction method using hidden neural networks" , Bioinformatics 21(2) ,pp 152-159, 2005.

[15] B. Rost, C. Sander, R. Schneider, PHD-an automatic mail server for protein secondary structure prediction, Bioinformatics 10(1) ,pp53-60. , 1994.

[16] C. Geourjon, G. Deleage, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, Bioinformatics 11(6), pp 681-684, 1995.