Biotechnology and Biological Sciences – Sen et al. (Eds)
© 2020 Taylor & Francis Group, London, ISBN 978-0-367-43161-7

Structure prediction and functional characterization of uncharacterized protein Rv1708 of *Mycobacterium tuberculosis* (Strain ATCC 25618/H37Rv)

Shamrat Kumar Paul, Tasnin Al Hasib, Pranta Ray, Lutful Kabir & Abul Bashar Ripon Khalipha*
Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Bangladesh

ABSTRACT: Tuberculosis (TB) is one of the most known ancient human disease which is caused by *Mycobacterium Tuberculosis* (MTB). Surprisingly, at present it is one of the major causes of death of human around the globe. Complete genome sequencing of MTB provides us a storehouse of genomic information to study about MTB's complex pathogenicity. Among 3924 Open Reading frame of MTB, the Uncharacterized Protein Rv1708 is encoded by Rv1708 gene and it is a inferred to be a cell cycle regulatory protein. It is anticipated that during cell division the Uncharacterized Protein Rv1708 play a critical role in septum formation of MTB. Thus, it becomes the mediator of cell cycle progression and cell division. Inhibition of bacterial cell division by blocking its associated protein is known to be therapeutic target for defeating disease. But the Uncharacterized Protein Rv1708 of the MTB remained unexplored. So, our aim is to propose the structural and functional features in addition to reveal the physicochemical properties. Homology modelling of the Uncharacterized Protein Rv1708 was generated by using Phyre2 and Swiss Model. The in-stability index generated by using the ExPasy's ProtParam tool shows that the Uncharacterized Protein Rv1708 is stable and its nature is acidic. Ramachandran map analysis by MolProbity reveals 95.5% of all residues were in allowed regions and 87.7% of all residues were in favored regions; which is indicating strong evidence of good quality of protein structure. This in-silico process will unveil the role of unexplored Uncharacterized Protein Rv1708 of MTB, and so it can pave the way for enriching our knowledge of the pathogenesis and drug-targeting approach for MTB.

1 INTRODUCTION

Mycobacterium tuberculosis (MTB), the causative agent of tuberculosis (TB), is one of the ancient bacterial species that still have a staggering impact on mortality rate since two million people die each year globally despite the global use of live, attenuated vaccine and antibiotics (Gengenbacher & Kaufmann, 2012). TB is caused by breathing in air contaminated with microscopic droplets that contain the untreated, active form of MTB. These microscopic droplets spread in the air when someone with an untreated, active form of TB speaks, coughs, sneezes, sighs, laughs, or spits (Malenki, 2018). Inhaling only a few droplets of such bacteria can lead to infected with TB and poses the lifetime risk of falling ill with it of 10 percent. In addition, people living with malnutrition, diabetes, HIV or people who use tobacco, possess a much higher risk of falling ill. Genome sequencing of MTB reveals that it is the second-largest bacterial genome sequence ever found which is rich in repetitive DNA, especially insertion sequences. It has also duplicated housekeeping genes and multigene families (Okou et al., 2007). Among 3,924 open reading frames of MTB, 91% of them have the potential of coding

*Corresponding author: khalipha1982@gmail.com

capacity. Some of them are belonging in-frame stop codons or frameshift mutations. This play crucial role in frameshifting during translation in association with pseudogenes (Cole et al., 1998). The genomic context of Uncharacterized Protein Rv1708 is observed and cross-checked by accession NC000962.3 from National Centre for Biotechnology Information (NCBI). Literature review showed that, a septal ring mediates the cell division of MTB, in association with a dozen of known proteins gather to the division site. In MTB, septal ring is a polymer of tubulin like Filamenting temperature-sensitive mutant Z (FtsZ) Protein. Septum formation and FtsZ polymerization are mediator of transcription and are clambered by protein interactions. A recent study shows that the Uncharacterized Protein Rv1708; a cell cycle regulatory protein that steers cell cycle progression and consequently cell division (Misra, Maurya, Chaudhary, & Misra, 2018). However, the structure of this Uncharacterized Protein Rv1708 is not reported yet. Also, the detailed physicochemical characterization and promising structure is not elucidated, so we have proposed a computer-aided homology modelled structure prediction of Uncharacterized Protein Rv1708 of MTB.

2 METHODOLOGY

2.1 *Retrieval of target amino acid sequence*

The amino acid sequence of Uncharacterized Protein Rv1708 of MTB (strain ATCC25618/ H37Rv) was retrieved from UniportKB database with the ID P9WLT1. Due to unavailability of 3-D structure in Research Collaboratory for Structural Bioinformatics (RCSB) protein data bank (PDB), modelling of this uncharacterized protein was undertaken utilizing 318 amino acid long sequence of Uncharacterized Protein Rv1708 of MTB (Consortium, 2019).

2.2 *Physicochemical characterization*

Physicochemical properties of the retrieved sequence were determined using two web-based servers, ExPasy's ProtParam tool used for the calculation and interpretation of amino acid composition, instability and aliphatic index, extinction coefficients and grand average of hydropathicity (GRAVY) (Gasteiger et al., 2009). Theoretical isoelectric point (pI) was also calculated using Sequence Manipulation Suite (SMS) Version2.

2.3 *Secondary structure prediction*

The self-optimized prediction method with alignment (SOPMA) (Geourjon & Deléage, 1995) and PSIPRED program was used to predict the secondary structure of Uncharacterized Protein Rv1708. Disorder prediction was performed using DISOPRED tool (Deng, Gumm, Karki, Eickholt, & Cheng, 2015).

2.4 *Homology modeling and validation*

There is no experimentally elucidation of 3D structure available for Uncharacterized Protein Rv1708 of MTB in RCSB protein data bank (PDB), therefore homology modelling of the protein of Uncharacterized Protein Rv1708 was done using two program Swiss Model (Schwede, Kopp, Guex, & Peitsch, 2003) and Phyre2 (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015). Secondary structure has also been predicted using Phyre2. 3D model of Uncharacterized Protein Rv1708 generated from Swiss Model and Phyre2 was compared and only the most suiTable 3D model was selected for final validation. The final modelled structure was validated using Ramachandran plot analysis by MolProbity for stereo-chemical property (Chen et al., 2010).

3 RESULT AND DISCUSSION

3.1 *Physicochemical characterization*

The amino acid sequence of Uncharacterized Protein Rv1708 of MTB was retrieved in FASTA format and used as query sequence for determination of physicochemical parameters. The instability index of Uncharacterized Protein Rv1708 of MTB is 31.61 (<40) indicated the stable nature of the protein (Guruprasad, Reddy, & Pandit, 1990). The molecular weight of protein is 34380.48Da and is acidic in nature (pI 6.00, 6.40*). (*pI determined by SMS Version2) High extinction coefficient values 16180 M^{-1} cm^{-1} indicate the presence of Cys, Trp and Tyr residues when all pairs of Cys residues are assumed to form cystines and the value is 15930 M^{-1} cm^{-1} when all Cys residues are reduced. The aliphatic index value (99.31) is pretty higher which suggested as a positive factor for increased thermos-stability in a wide range of temperature The nature of protein found hydrophilic which indicate the possibility of better interaction with water (Ikai, 1980) and it was calculated by the lower grand average of hydropathicity (GRAVY) indices value (-0.070) as shown in Table 1.

3.2 *Secondary structure prediction*

For the prediction of secondary structure by SOPMA the parameters (window width: 17; similarity threshold: 8; division factor: 4) were considered as default. Analyzing 511 proteins (sub-database) and 33 aligned proteins, SOPMA predicted 39.62 percent of residues as random coils in comparison to alpha helix (37.74 percent), extended strand (17.92 percent) and Beta turn (4.72 percent) as shown in Table 2. PSIPRED shows the higher confidence of prediction of helix, strand and coil. Intrinsic disorder profile was computed using DISOPRED and <87 percent of the amino acid are below the confidence score of 0.5 for disordered condition, suggested the lowest possibility of distortion and conferred the high stability to the predicted protein (Figure 2 Left).

3.3 *Homology modelling and structural validation*

FASTA format sequence of Uncharacterized Protein Rv1708 was inserted as target sequence in Swiss-Model workspace. Both HHBlits and BLAST parallelly was used to search the Swiss Model Template Library (SMTL). Total 6327 templates were found to match the target sequence resulted in total 50 templates, among the five most favorable template showed in Table 3 (2bek.1.A, 2bek.1.A, 2bej.1.A, 2bej.1.A, 1wcv.1.A) are those who have GMQE value greater than 0.50. Target sequence was selected based on the Qualitative Model Energy Analysis

Table 1. Physicochemical parameters computed using Expasy's Prot-Param and SMS tool.

Physio-chemical parameters	Values
No. of amino acid (aa)	318
Molecular weight (MW)	34380.48
Theoretical Isoelectric point (pI)	6.00, 6.40*
Aliphatic Index	99.31
Instability Index	31.61
Extinction Coefficient (All Cys form Cysteine)	16180
Extinction Coefficient (All Cys reduced)	15930
Total no. of negatively charged residues (Asp + Glu)	38
Total no. of positively charged residues (Arg + Lys)	34
GRAVY (Grand average of hydropathicity)	-0.070

* pI determined by SMS Version2

Table 2. Secondary structure elements prediction by SOPMA.

Secondary structure elements	Values (%)
Alpha helix (Hh)	37.74%
3_{10} helix (Gg)	0.00%
Pi helix (Ii)	0.00%
Beta bridge (Bb)	0.00%
Extended strand (Ee)	17.92%
Beta turn (Tt)	4.72%
Bend region (Ss)	0.00%
Random coil (Cc)	39.62%
Ambiguous states (?)	0.00%

Table 3. Alignment of selected template.

Name	Title	GMQE	QSQE	Identity (%)
2bek.1.A	Segregation protein	0.58	0.53	43.21
2bek.1.A	Segregation protein	0.60	0.51	46.61
2bej.1.A	Segregation protein	0.57	0.35	43.62
2bej.1.A	Segregation protein	0.59	0.33	47.03
1wcv.1.A	Segregation protein	0.57	0.34	43.62

(QMEAN) score (-1.85), Global model quality estimate (GMQE) 0.58, percentage of sequence identity (43.21 percent), similarity (0.40) and coverage (0.76). Model was based on target-template alignment using Swiss model workbench where insertion, deletions remodeled and side chains were then rebuilt. Our model showed resemblance with 2bek.1.A; which is the bacterial chromosome segregation protein Soj of *Thermus thermophilus* and identified as the template for Uncharacterized Protein Rv1708 homology modeled structure. So, the model generated was by using the template 2bek.1.A of Segregation Protein considering the and saved in PDB format and visualized by Discovery Studio 4.5 Visualizer (Figure 1 left). The respective values Z-scores of CBeta interaction energy, torsion angle energy, solvation energy, secondary structure in case of Uncharacterized Protein Rv1708 are -1.94, -1.49, 0.18 and -0.56. The overall QMEAN score for Uncharacterized Protein Rv1708 is -1.85. By QMEAN generated results of Uncharacterized Protein Rv1708, it is conferred as a qualified model for drug target scopes. Similarly, the

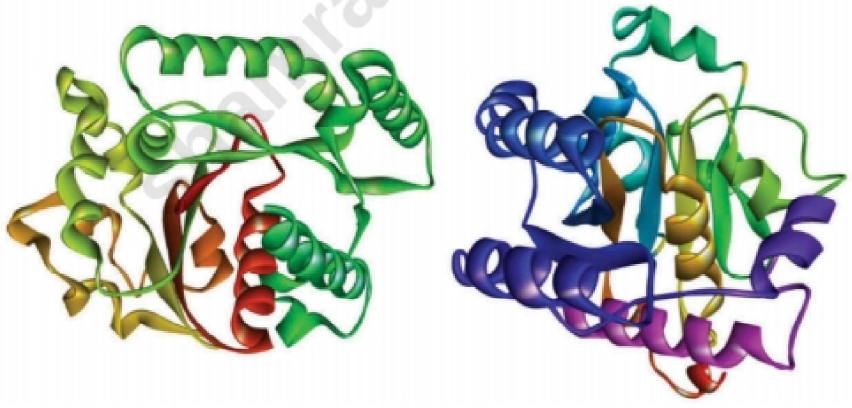

Figure 1. Uncharacterized Protein Rv1708 Structure by Swiss Model (Left) & Phyre2 (Right).

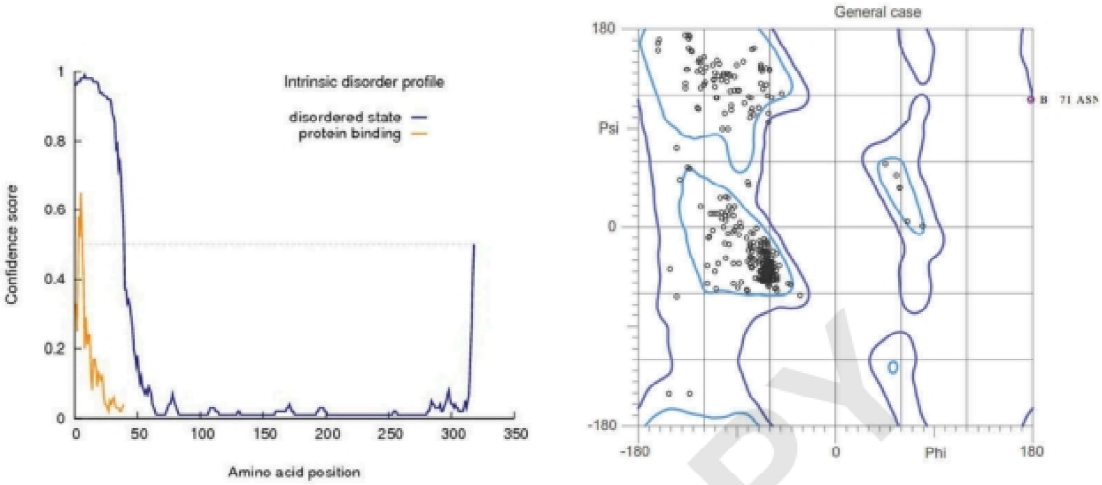

Figure 2. Intrinsic Disorder Profile (Left), Ramachandran Plot using MolProbity (Right).

Table 4. Ramachandran plot analysis by MolProbity.

Ramachandran plot statistics	Residue	%
Residues in the favored regions	481	95.4
Residues in the allowed regions	502	99.6
Residues in the outliers (phi, psi)	02	0.40
Total number of residues	504	-

homology modelling of Uncharacterized Protein Rv1708 was performed by Phyre2. Based on the 6 templates (c3ez6B, c2ozeA, d2afhe1, c2bekB, crpfsA, c5aorA and c3q9lB), protein model was generated with 85 percent of the residues modelled at 100 percent confidence (Figure 1-right). Both predicted structural models were evaluated by MolProbity for stereochemical property. Between two structures, Ramachandran Plot data (Figure 2 right) from Swiss Model is promising and that's why we have considered structure from Swiss Model for the result analysis. Here 95.4% of all residues were in favored (98%) regions and 99.6% of all residues were in allowed (>99.8%) regions. There were 2 outliers (phi, psi): Asn71 Val131 and these data of predicted structure from Swiss Model ensures the good quality of the protein structure; data showed in Table 4.

4 CONCLUSION

In this study, we have concluded the structural model of Uncharacterized Protein Rv1708 (strain ATCC 25618/H37Rv) of MTB through in-silico approach. The physicochemical parameters prediction and functional annotation are useful for understanding the action of this proteins' activity. Our Homology-Modelled protein provides insights into the functional role of Uncharacterized Protein Rv1708 (strain ATCC 25618/H37Rv) in pathogenesis, which will help to design potential therapeutic drug against this protein for inhibition of septum formation of MTB.

ACKNOWLEDGMENT

We are grateful to the Department of Biochemistry and Molecular Biology of Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj-8100, Bangladesh for giving us a computational platform for completing this project.

REFERENCES

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Richardson, D.C. (2010). MolProbity : all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography*. https://doi.org/10.1107/s0907444909042073

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Barrell, B.G. (1998). Deciphering the biology of mycobacterium tuberculosis from the complete genome sequence. *Nature*. https://doi.org/10.1038/31159

Consortium, T. U. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky1049

Deng, X., Gumm, J., Karki, S., Eickholt, J., & Cheng, J. (2015). An overview of practical applications of protein disorder prediction and drive for faster, more accurate predictions. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms160715384

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., & Bairoch, A. (2009). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*. https://doi.org/10.1385/1-59259-890-0:571

Gengenbacher, M., & Kaufmann, S.H.E. (2012). Mycobacterium tuberculosis: Success through dormancy. *FEMS Microbiology Reviews*. https://doi.org/10.1111/j.1574-6976.2012.00331.x

Geourjon, C., & Deléage, G. (1995). Sopma: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/11.6.681

Guruprasad, K., Reddy, B.V.B., & Pandit, M.W. (1990). Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering, Design and Selection*, *4*(2), 155–161. https://doi.org/10.1093/protein/4.2.155

Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *Journal of Biochemistry*.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., & Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*. https://doi.org/10.1038/nprot.2015.053

Malenki, G. (2018). The use of isoniazid as prophylaxis in patients co-infected with tuberculosis and HIV. *Canadian Journal of Respiratory Therapy*.

Misra, H.S., Maurya, G.K., Chaudhary, R., & Misra, C.S. (2018). Interdependence of bacterial cell division and genome segregation and its potential in drug development. *Microbiological Research*. https://doi.org/10.1016/j.micres.2017.12.013

Okou, D.T., Steinberg, K.M., Middle, C., Cutler, D.J., Albert, T.J., & Zwick, M. E. (2007). Microarray-based genomic selection for high-throughput resequencing. *Nature Methods*. https://doi.org/10.1038/nmeth1109

Schwede, T., Kopp, J., Guex, N., & Peitsch, M.C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Research*.