Structural and Functional Annotation of Uncharacterized Protein NCGM946K2_146 of Mycobacterium tuberculosis: An In-Silico Approach

Abu Saim Mohammad Saikat[1], Rabiul Islam[2]*, Shahriar Mahmud[3], Md. Abu Sayeed Imran[3],Mohammad Shah Alam[4], and Md. Ekhlas Uddin[5]

[1]Dept. of Biochemistry and Molecular Biology, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh; [2]Divisional DNA Screening Laboratory, Faridpur Medical College Hospital, Ministry of Women & Children Affairs, Dhaka, Bangladesh; [3]Dept. of Biotechnology and Genetic Engineering, Islamic University, Kushtia, Bangladesh; [4]Dept. of Microbiology, Gono Bishwabidalay, Savar, Dhaka, Bangladesh; and [5]Dept. of Biochemistry and Molecular Biology, Gono Bishwabidalay, Savar, Dhaka, Bangladesh.

ABSTRACT

The human pathogen of *Mycobacterium tuberculosis* (MTB) is indeed one of the renowned important longtime infectious diseases, tuberculosis (TB). Interestingly, MTB infection has become one of the world's leading causes of human death. In trehalose synthase, the protein NCGM 946K2 146 found in MTB has an important role. For carbohydrate transport and metabolism, trehalose synthase is required. The protein isn't clarified yet, though. In this research, an *in silico* approach was therefore formulated for functional and structural documentation of the uncharacterized protein NCGM946K2 146. Three distinct servers, including Modeller, Phyre2, and Swiss Model were used to evaluate the predicted tertiary structure. The top materials are selected using structural evaluations conducted with the analysis of Ramachandran Plot, Swiss-Model Interactive Workplace, Prosa-web, Verify 3D, and Z scores. This analysis aimed to uncover the value of the NCGM946K2 146 protein of MTB. This research will, therefore, improve our pathogenesis awareness and give us a chance to target the protein compound.

INTRODUCTION

Mycobacterium tuberculosis (MTB) is an antiquity bacterial species which is a rod-like, acid-fast, and Gram-positive organism responsible for one of the most lethal diseases (ranking above HIV/AIDS) – tuberculosis (TB). Typically, TB spreads from an individual infected with MTB via the air, such as by coughing. Pulmonary TB is the infection of the lungs, and extrapulmonary TB is the infection of other sites of the body. It has been reported in 2018 that almost 10 million tuberculosis patients (range 9.0–11.1 million) died from HIV-negative deaths in 2018, creating a high risk for tuberculosis spreading and global growth [1]

MBT has the second-largest bacterial genome sequence on tap with 3924 open reading frames. Indeed, multi-gene families and duplicate housekeeping genes have multiple, repetitive DNAs, particularly insertion sequences present in MTB [2]. The CD-search tool [3] predicted a domain of the protein NCGM946K2_146 and was described as a functional protein. Moreover, the uncharacterized protein NCGM946K2_146 from MTB is structurally, and functionally is not reported. The analysis, therefore, explains the comprehensive physicochemical characterization and the predicted functionally annotated tertiary structure.

MATERIALS AND METHODS

Sequence retrieval

The amino acid sequence of NCGM946K2_146 was retrieved in FASTA format from the National Center for Biotechnology Information (NCBI) [4] with the accession ID of BAW10952. Yet, up to this point, NCGM946K2 146 is not accessible in the Protein Data Bank (PDB) as a tertiary structure of the unknown protein. The structural patterns of this protein subsequently began using the protein NCGM946K2 146 with a 455 amino acid long chain.

Physicochemical Characterization

The ExPASy server ProtParam method has been used to measure the amino acid sequence composition, the instability index, the aliphatic index, the GRAVY, and extinction coefficients as well [5].Moreover, the SMS Suite (v2.0) for the measurement of the theoretical isoelectricpoint (pI) of the NCGM946K2 146 protein was performed[6].

MATERIALS AND METHODS

Functional Annotation Prediction

The CD Search tool of NCBI [3] is used for domain prediction. The CD Search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) predicted a domain of the protein NCGM946K2_146.

Secondary Structure Prediction

The secondary structure NCGM946K2_146has been predicted by the SPIPRED software (server-based) and the SOPMA framework was used for elements prediction[7] [8].

Tertiary Structure Modeling and Validation

Currently, there is no experimentally concluded tertiary structure available for NCGM946K2_146 of MTB in the Protein Data Bank (PDB). Consequently, the tertiary structures of the protein modeled by utilizing three different programs, including Modeller [9] with the HHpred tool [10], Phyre2 [11], and Swiss-Model server [12].

MATERIALS AND METHODS

Sub-cellular localization

The subcellular localization predictor tool **CELLO v. 2.5** (http://cello.life.nctu.edu.tw/), executed for sub-cellular localization of the protein NCGM946K2_146 present in MTB. This tool is performed for the amino acid comp., N-peptide comp., physicochemical comp., neighboring seq. Comp. and partitioned seq. Comp. of the protein.

RESULTS AND DISCUSSION

Secondary Structure Elements	Values (%)
Alpha helix (Hh)	51.21
3_{10} helix (Gg)	0.00
Pi helix (Ii)	0.00
Beta bridge (Bb)	0.00
Extended strand (Ee)	11.21
Beta turn (Tt)	3.74
Bend region (Ss	0.00
Random coil (Cc)	33.85
Ambiguous states	0.00
Other states	0.00

Table 2: Secondary Structure Elements.

Servers	Ramachandran Plot Calculation	Value (%)
Modeller	Residues in most favored regions [A,B,L]	95.2
	Residues in additional allowed regions [a,b,l,p]	4.8
	Residues in generously allowed regions [~a,~b,~l,~p]	0.0
	Residues in disallowed regions	0.0
Phyre2	Residues in most favored regions [A,B,L]	93.6
	Residues in additional allowed regions [a,b,l,p]	6.1
	Residues in generously allowed regions [~a,~b,~l,~p]	0.0
	Residues in disallowed regions	0.3
Swiss Model	Residues in most favored regions [A,B,L]	94.2
	Residues in additional allowed regions [a,b,l,p]	5.3
	Residues in generously allowed regions [~a,~b,~l,~p]	0.4
	Residues in disallowed regions	0.1

Table 5: Ramachandran Plot Analysis.

RESULTS AND DISCUSSION

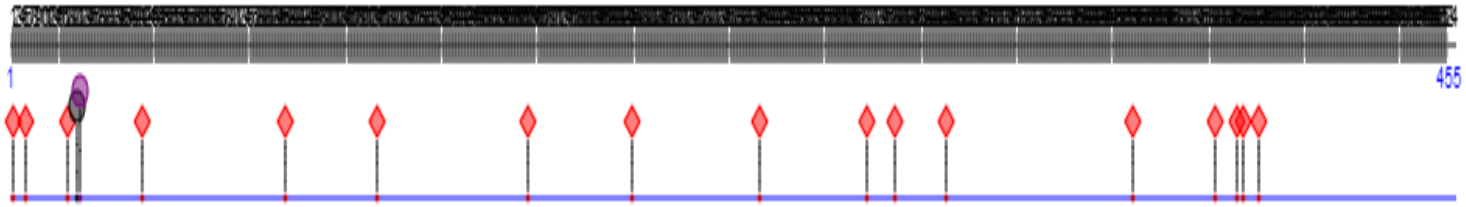

Figure 2: Protein-Protein and Protein-Polynucleotide Binding Sites. It showed there were **17 different active protein binding sites** including the position of viz: **1-2; 5; 18-19; 42-44; 87-88; 116; 164; 197-199; 237-238; 271-273; 280; 296; 355; 381-383, 388; 390; and 395.**

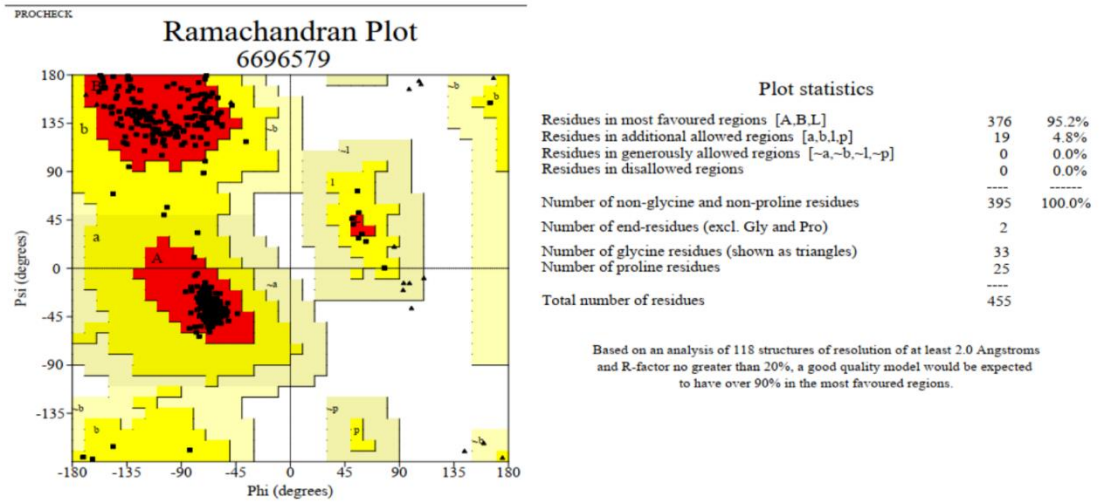

Figure 5: Ramachandran Plot Analysis of Modeller Predicted Protein Structure

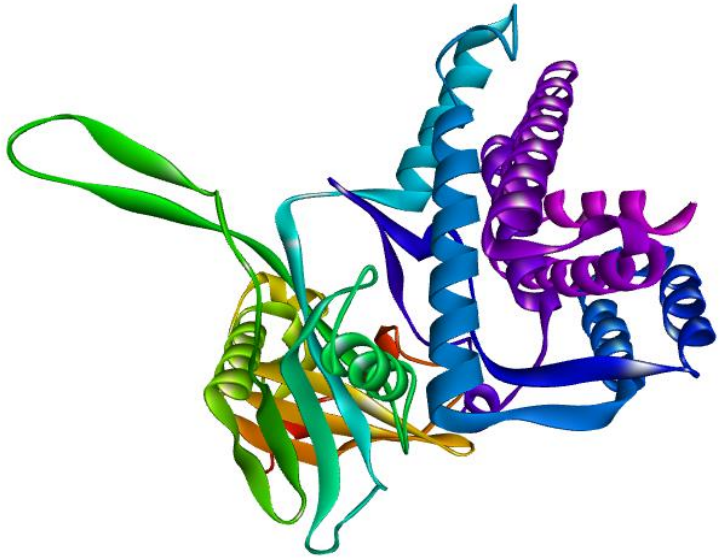

Figure 3: Structure of NCGM946K2_146 Predicted by Modeller.

CONCLUSION

The structural as well as the functional annotation of NCGM946K2_146, which is located in MTB, was documented in this study with the predicted ligand-binding active sites present in *M. tuberculosis*. The arrangement of amino acid sequences in the desired region was determined by assessing the protein structure. Regarding understanding protein operations, the physicochemical parameters also functional enrichment estimation is beneficial. The secondary assumption and evaluation structures verified that alpha-helix, random spiral, extended strand, and beta turns were predominant in most sequences. Three different servers including the Modeller, the Phyre2, and the Swiss Model servers have assessed the assumed tertiary structures. PROCHECK for Ramachandran Map Analysis, the Verify 3D, the Swiss-Model Interactive Workplace server, and the Z-scores from Prosa-Web used as protein structure evaluation tools. The results showed that the Modeller is appropriate *in silico* documentation for the modeled protein NCGM946K2_146 from the three separate servers. This study would provide an opportunity to design effective therapeutic drugs against the protein of *M. tuberculosis*.

REFERENCES

1. WHO | Global tuberculosis report 2019 (2019) URL https://www.who.int/tb/publications/global_report/en/
2. Yamurai Bishi L, Chaitanya Vedithi S, L. Blundell T, Mugumbate G. Computational Deorphaning of Mycobacterium tuberculosis Targets. Drug Discov. Dev. - New Adv., IntechOpen; 2019. https://doi.org/10.5772/intechopen.82374
3. CDD/SPARCLE: the conserved domain database in 2020. Nucleic acids research 48, 48(D1), D265-D268.


THANK
YOU