P: ISSN NO.: 2394-0344 RNI No.UPBIL/2016/67980 VOL-5* ISSUE-8* November- 2020
E: ISSN NO.: 2455-0817 *Remarking An Analisation*

Bioinformatic Analysis of Physiochemical Properties of Rv3334 Protein from Mycobacterium Tuberculosis

Paper Submission: 10/11/2020, Date of Acceptance: 20/11/2020, Date of Publication: 21/11/2020

Bhanu Pratap
Research Scholar
Dept. of School of Life Sciences,
Jawaharlal Nehru University
New Delhi, India

Manoj Kumar
Research Scholar
Dept. of School of Life Sciences,
Jawaharlal Nehru University
New Delhi, India

Kumar Sanbhav Singh
Research Scholar
Dept. of Botany
B.R.A. Bihar University,
Muzaffarpur, Bihar

Abstract

Mycobacterium tuberculosis can persist in the host without showing any symptoms. It survives under hypoxic conditions by regulating a set of genes. Rv3334 protein from Mycobacterium tuberculosis is upregulated under hypoxic conditions. It is a member of the MerR family of transcriptional regulators. In the current study, I have analyzed the physiochemical properties of Rv3334 protein using the bioinformatics tool. Analysis of the primary structure and secondary structure of Rv3334 was done. Based on the analysis, Rv3334 is a stable protein and rich in negatively charged amino acids having a molecular weight of 16169.24 Dalton. Domain present in the Rv3334 was identified by using the Pfam tool. It was found that protein contains two distinct domains, N-terminal domain sensing domain, and C-terminalDNA binding domain. The Secondary structure analysis was done using DSC, RaptorX, HNN, PHD, and SOPMA online server and it was found that protein is rich in α-helical content.

Keywords: Mycobacterium tuberculosis, Transcriptional regulator, Dormancy, Virulence, Pathogen

Introduction

Tuberculosis is caused by Mycobacterium tuberculosisbacterium. It is one of the successful human pathogens. According to the World Health Organization (WHO), 30% of the world's population is infected with Mycobacterium tuberculosis. The majority of infected individuals carry Mycobacterium tuberculosis without showing any symptoms. It can persist for decades in the host cell and it is termed as latent tuberculosis infection (LTBI) (Stewart, Robertson, and Young; North and Jung)This condition is called a dormant stage. Complete eradication of tuberculosis is tough due to the dormant stage. The mechanism of dormancy and persistence is still unknown. Mycobacterium tuberculosis has a set of genes which helps in the survival during stress condition like hypoxia, limited availability of nutrients, acidic pH, high temperature, and oxidative and nitrosative stress inside the host. PhoPR a two-component system of Mycobacterium tuberculosis regulates many genes that are involved in hypoxia. An earlier study showed that Rv3334 is upregulated during oxidative, nitrosative, and hypoxia stress conditions (Muttucumaru et al.; Voskuil et al.). Rv3334 undergoes autorepression and upregulates the KstR protein (Gomez et al.).

It is also upregulated by Rv0081, a hypoxic transcription regulator(Sun et al.). A detailed study of Rv3334 and its mechanism of gene regulation have not yet been done. In this study, I have analyzed the primary structure and secondary structure of Rv3334. A bioinformatics tool was used to study the physiochemical properties of the protein. Based on the analysis, Rv3334 is a stable protein and rich in negatively charged amino acids. Secondary structure analysis shows that protein is rich inα-helical content.

Objectives of study

The main objectives of this study are

1. To understand the physiochemical properties of the Rv3334 protein.
2. To understand the primary and secondary structure of the protein

Review of literature

Tuberculosis (TB) has remained one of the major global health problems taken all sort of dimension, particularly in developing countries. Mycobacterium tuberculosis is the pathogen that causes tuberculosis. 30% of the population is infected with Mycobacterium tuberculosis and it causes three billion deaths every year(WHO). It was found that three closely related species of mycobacteria, namely Mycobacterium bovis, Mycobacterium africanum, and Mycobacterium tuberculosis have been recognized as the aetiological agents of tuberculosis in humans. M. tuberculosis is spread through aerosol and inhaled bacilli are engulfed by alveolar macrophages in the lungs. The virulence of these pathogens is due to their ability to parasitize and survive within phagocytic cells(Fenton and Vermeulen). Latent tuberculosis infection (LTBI) can persist in the host cell for decades without showing any symptoms, the phenomenon is called dormancy (Stewart, Robertson, and Young). It is thought that adaptation to hypoxia play important role in persistence but the mechanism is unknown. In Mycobacterium tuberculosis, Rv0081 and Rv3334 connect the early and enduring hypoxic response(Sun et al.). According to Sun.et.al, Rv0081 and Rv3334 belong to the same regulatory pathway. The mutant form of Rv0081 and Rv3334 show defective growth in the hypoxic condition.Rv0081 binds directly to the promotor of Rv3334 and upregulates it.

Methodology

Sequence retrieval and primary structure

The amino acid sequence of Rv3334 from Mycobacterium tuberculosis H37Rv (UniProt id: O53384) was retrieved from the UniProtKB database to analyze their physiochemical and 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)(Marchler-Bauer et al.). The primary structure and the amino acid composition of Rv3334 were computed using Protparamtools(Gasteiger et al.). The presence of domain was identified using the Pfam tool(El-Gebali et al.). Amino acid sequence retrieved from the UniProtKB database was used as the input sequence.

Secondary structure prediction

The secondary structure of Rv3334 was predicted by using online tools like DSC(King and Sternberg), RaptorX(Wang et al.), HNN(Lin et al.), PHD(Rost, Sander, and Schneider), and SOPMA(Geourjon and Deleage) the secondary amount was calculated.

Physiochemical properties

The amino acids of Rv3334 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 ProtParmtool(Gasteiger et al.).

Results and Discussion

Primary structure

Rv3334 gene is located between two conserved genes Rv3333c and Rv3335c in the Mycobacterium tuberculosis genome. The conserved domain database (CDD) analysis shows that this protein belongs to a member of MerR transcriptional regulators. Pfam analysis shows that Rv3334 contains two distinct domains. N-terminal domain and C-terminal contains Helix-Turn-Helix (HTH) DNA binding domain. The gene sequence of Rv3334 is as follows.

>NC_000962.3:3721257-3721697 Mycobacterium tuberculosis H37Rv, complete genome

ATGAAGATCAGCGAGGTAGCCGCGCTCACCAACA
CCAGCACCAAGACCCTCCGCTTCTACGAGAACTC
GG
GGCTGCTGCCGCCGCCTGCACGCACAGCATCGG
GGTATCGCAACTATGGACCCGAGATCGTGGATCG
GCT
GCGGTTTATCCATCGGGGCCAAGCGGCCGGGCT
GGCATTACAGGAAGTACGCCAAATCCTGGCCATC
CAC
GACCGCGGCGAGGCGCCGTGCGCACACGTCCG
CCAACTACTGAGCACCCGCATCGACGAAGTCCGC
GCGC
AGATCGCCGAACTGATTGCCCTCGAAGGCCACTT
GCAGACCCTGCTTGACCACGCTTCATATGGCCCG
CC
CACCGAACACGACCACTCCACGGTGTGTTGGATC
CTGGAAAGCGACCTCGATGAGCCCACCGCCATC
GAG
GTCAGCGACATTCACGCCTAG

The Rv3334 is a 146 amino acid long protein (Met1to Ala146). The amino acid sequence was retrieved from UniprotKB which is as under here. Details of the composition of amino acids are mentioned in Table 1. Atomic composition calculated using the protparam tool is mentioned in Table 2. Alanine, Isoleucine, glutamate, arginine, threonine, and leucine are predominantly present in the protein. Details of amino acid composition are as follows.

>NP_217851.1 MerR family transcriptional regulator [Mycobacterium tuberculosis H37Rv]

MKISEVAALTNTSTKTLRFYENSGLLPPPARTASGY
RNYGPEIVDRLRFIHRGQAAGLALQEVRQILAIH
DRGEAPCAHVRQLLSTRIDEVRAQIAELIALEGHLQT
LLDHASYGPPTEHDHSTVCWILESDLDEPTAIE
VSDIHA

Table 1- Composition of amino acids

Name of Amino acid	No. of molecule present	Percentage
Ala (A)	16	11.0%
Arg (R)	11	7.5%
Asn (N)	3	2.1%
Asp (D)	8	5.5%
Cys (C)	2	1.4%
Gln (Q)	6	4.1%
Glu (E)	12	8.2%
Gly (G)	8	5.5%
His (H)	8	5.5%
Ile (I)	11	7.5%
Leu (L)	17	11.6%
Lys (K)	2	1.4%

Met (M)	1	0.7%
Phe (F)	2	1.4%
Pro (P)	8	5.5%
Ser (S)	9	6.2%
Thr (T)	10	6.8%
Trp (W)	1	0.7%
Tyr (Y)	4	2.7%
Val (V)	7	4.8%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

Table 2- Atomic composition of amino acids

Name of atom	No. of atoms present
Carbon	710
Hydrogen	1133
Nitrogen	207
Oxygen	219
Sulfur	3

Physiochemical properties

Physiochemical characterization is very important to characterize specific proteins. The average Molecular weight of Rv3334 proteins calculated is 16169.24 dalton. ProtParam tool was used to study the physiochemical properties of Rv3334 proteins. The results are shown in Table 3. Protein pI is calculated using the pKa values of amino acids.

Table 3. Physiochemical Properties of Rv3334

Properties	Value
No. of Amino acids	146
Mol. Wt. in Dalton	16169.24
Theoretical pI	5.54
Asp+Glu	20
Arg+Lys	13
Extinction coefficients (EC)	11460
Instability index	38.03
Aliphatic index	99.66
GRAVY	-0.246

Secondary structure prediction

The secondary structure of Rv3334 protein was predicted using online tools like DSC, RaptorX, HNN, PHD, and SOPMA. The percentage of α-Helix, β-Sheets, and Random coil was calculated and it was found that Rv3334 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 Rv3334

Tools	α-Helix (%)	β-Sheets (%)	Random coil (%)
DSC	63.4	0.0	36.5
RaptorX	58.0	3.0	37.0
HNN	45.5	11.7	42.7
PHD	55.8	2.0	41.0
SOPMA	62.0	8.0	35.9

Conclusion

Rv3334 is a two-domain protein having a molecular weight of 16169.24 Dalton protein.Primary structure analysis shows that Rv3334 is a negatively charged protein due to the presence of the high content of (Asp + Glu: 20). The aliphatic index (99.66) and the instability index are 38.03 which classified protein as stable. The Secondary structure analysis shows that Rv3334 proteins have predominant α-helical structures.

References

1. *El-Gebali, Sara, et al. "The Pfam Protein Families Database in 2019." Nucleic acids research 47 D1 (2019): D427-D32.*
2. *Gasteiger, Elisabeth, et al. "Protein Identification and Analysis Tools on the Expasy Server." The Proteomics Protocols Handbook. Springer, 2005. 571-607.*
3. *Fenton, Matthew J, and Mary W Vermeulen. "Immunopathology of Tuberculosis: Roles of Macrophages and Monocytes." Infection and immunity 64 3 (1996): 683.*
4. *Geourjon, Christophe, and Gilbert Deleage. "Sopma: Significant Improvements in Protein Secondary Structure Prediction by Consensus Prediction from Multiple Alignments." Bioinformatics 11 6 (1995): 681-84.*
5. *Gomez, Roshna Lawrence, et al. "The Multiple Stress Responsive Transcriptional Regulator Rv3334 of Mycobacterium Tuberculosis Is an Autorepressor and a Positive Regulator of Kstr." The FEBS Journal 283 16 (2016): 3056-71.*
6. *King, Ross D, and Michael JE Sternberg. "Identification and Application of the Concepts Important for Accurate and Reliable Protein Secondary Structure Prediction." Protein science 5 11 (1996): 2298-310.*
7. *Lin, Kuang, et al. "A Simple and Fast Secondary Structure Prediction Method Using Hidden Neural Networks." Bioinformatics 21 2 (2005): 152-59.*
8. *Marchler-Bauer, Aron, et al. "Cdd: A Conserved Domain Database for the Functional Annotation of Proteins." Nucleic acids research 39 suppl_1 (2010): D225-D29.*
9. *Muttucumaru, DG Niranjala, et al. "Gene Expression Profile of Mycobacterium Tuberculosis in a Non-Replicating State." Tuberculosis 84 3-4 (2004): 239-46.*
10. *North, Robert J, and Yu-Jin Jung. "Immunity to Tuberculosis." Annu. Rev. Immunol. 22 (2004): 599-623.*
11. *World HealthOrganization. "Global Tuberculosis Report 2018. 2018." Geneva: World Health Organization (2019)*
12. *Rost, Burkhard, Chris Sander, and Reinhard Schneider. "Phd-an Automatic Mail Server for Protein Secondary Structure Prediction." Bioinformatics 10 1 (1994): 53-60.*
13. *Stewart, Graham R, Brian D Robertson, and Douglas B Young. "Tuberculosis: A Problem with Persistence." Nature Reviews Microbiology 1 2 (2003): 97-105.*
14. *Sun, Xian, et al. "Transcription Factors Rv0081 and Rv3334 Connect the Early and the Enduring Hypoxic Response of Mycobacterium Tuberculosis." Virulence 9 1 (2018): 1468-82.*
15. *Voskuil, Martin I, et al. "The Response of Mycobacterium Tuberculosis to Reactive Oxygen and Nitrogen Species." Frontiers in microbiology 2 (2011): 105.*
16. *Wang, Sheng, et al. "Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields." Scientific reports 6 1 (2016): 1-11*