Journal of Veterinary & Marine Sciences

JVMS, 2(2): 88-93 www.scitcentral.com



Original Research Article: Open Access

Computational Analysis of Mycobacterium bovis Protein Sequence

Dauda A^{1*}, Abbaya HY² and Ndirpaya AD³

*1 Department of Animal Science, University of Calabar, P.M.B. 1115 Calabar, Nigeria

²Department of Animal Production, Adamawa State University (ADSU), Mubi, P.M. B. 025, Adamawa State, Nigeria

³Department of Microbiology, University of Maiduguri, P.M.B 1069 Maiduguri, Nigeria.

Received October 16th, 2019; Revised November 09th, 2019; Accepted November 11th, 2019

ABSTRACT

A total of ten (10) *Mycobacterium bovis* (MBB) proteins were retrieved from the GenBank. The GenBank accession numbers of the sequences and sequence variations of the proteins were used to investigate the molecular identity of various MBB proteins. The physico-chemical properties of MBB proteins were performed using ExPASy Protparam tool. The amino acid sequences of MBB proteins were subjected to secondary structure prediction using ExPASy's SOPMA tool. The three dimensional structure (3D) of the MBB protein were determined using Phyre2 server. The 3D structure was then validated by using VADAR server and Ramachandran plot. Isoelectric point (pI), molecular weight (MW), extinction coefficient (EC); instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY) were computed. The study revealed that some of the MBB proteins were acidic while some were basic in nature based on their pI. The EC of all the proteins have values >100 while others were <100. The GRAVY of MBB proteins revealed some positive and negative values. The amino acid composition of MBB proteins indicates that they are rich in aliphatic amino acids. Among the MBB proteins some are alpha helice while others are random coil structure. The Ramachandran Plot validated the 3D structure of MBB proteins. The present protein information may aid future research in mutagenesis and pharmacogenetics of cattle especially in a developing economy such as Nigeria.

Keywords: Protein, Mycobacterium bovis, Sequences, Bioinformatics

INTRODUCTION

Mycobacterium bovis is the classical causative agent of bovine tuberculosis, can be responsible for human tuberculosis, which makes this bacterium an important zoonotic species. The Mycobacterium tuberculosis complex (MTBC) includes *M. tuberculosis* (the cause of most human tuberculosis), M. bovis, M. bovis bacillus Calmette-Guérin, M. africanum and M. microti [1]. M. bovis is the main cause of tuberculosis in cattle, deer and other mammals. The human bacillus M. tuberculosis may have evolved from M. bovis in the setting of animal domestication [2]. Human M. bovis infection generally occurs in the setting of consumption of infected cow milk products. The combination of disease tracing and molecular typing is needed to understand the epidemiology of tuberculosis. Mycobacterium bovis in silico study is significant because it help to bring new insights into epidemiological questions [3] Molecular typing has been instrumental in determining the population structure and evolution of pathogens. Since tuberculosis have economical and nutritional consequences, efforts should be intensified towards finding sustainable genomic solutions to these deadly diseases which continue to ravage the livestock industry [3]. New typing tool such as computational approach may help to improve the surveillance and control of the disease, as well as to trace new epidemics. The aim of the study is to carry out *in silico* study on the proteins sequence of *Mycobacterium bovis*.

MATERIALS AND METHODS

A total of ten (10) *Mycobacterium bovis* (MBB) nucleotide sequences were retrieved from the GenBank (NCBI) via (http://www.ncbi.nlm.nih.gov). The Genbank accession

Corresponding author: A Dauda, Department of Animal Science, University of Calabar, P.M.B. 1115 Calabar, Nigeria, Tel: (+234)7036707201; E-mail: ayubadauda87@gmail.com

Citation: Dauda A, Abbaya HY & Ndirpaya AD. (2020) Computational Analysis of *Mycobacterium bovis* Protein Sequence. J Vet Marine Sci, 2(2): 88-93.

Copyright: ©2020 Dauda A, Abbaya HY & Ndirpaya AD. This is an openaccess article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. numbers of the sequences are AAD09878, 3I7J, BAA05497, ANG90842, BAA05496, CAA44268, ANG90839. ANG90831. ANG90823 and ANG90818.ExPASv ProtParam Tool was used for the computation of various physical and chemical properties of the MBB proteins using amino acid sequences. The computed parameters were Isoelectric point (pI), molecular weight (MW), extinction coefficient (EC); instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY) [4]. The amino acid sequences of MBB proteins were subjected to secondary structure prediction using ExPASy's SOPMA tool as it is an improved SOPM method. It predicts 69.5% of amino acids for a 3 state description of the secondary structure (a helix, b sheets and coil). The Phyre2 server was used to predict the 3D structure of MBB proteins. These servers predict the three-dimensional structure of a protein sequence using the principles and techniques of homology modeling [5]. Currently, the most powerful and accurate methods for detecting and aligning remotely related sequences rely on profiles or Hidden Markov Models (HMMs). 3D ligand site was used to predict the binding site of the 3D structure of the MBB proteins. Phyre2 is coupled to the 3D ligand site server for protein binding site prediction [6]. The 3D structure was then validated by using VADAR server and Ramachandran plot. VADAR (Volume, Area, Dihedral Angle Reporter) is a compilation of more than 15 different algorithms and programs for analyzing and assessing peptide and protein structures from their PDB coordinate data. The results have been validated through extensive comparison to published data and careful visual inspection. The VADAR web server supports the submission of either PDB formatted files or PDB accession numbers. VADAR produces extensive tables and high quality graphs for quantitatively and qualitatively assessing protein structures determined by X-ray crystallography, NMR spectroscopy, 3D threading or homology modeling [7].

RESULTS

The physical and chemical properties of MBB protein are shown in Table 1. There were variations in the amino acid number of MBB proteins. The higher the amino acid number, the higher the molecular weight of the protein. The isoelectric point (pI) of the protein indicates that some were acidic with pI value less than 7 while some wee basic with pI value greater than 7. The net charges of some proteins were negative while others were positive. The EC value ranges from 30940-147140. The half-life of all the proteins indicates 30 h except that of protein with accession number 3I7 J which was not determined due to the presence of an Nterminal ambiguity. The instability index (II) of the proteins shows only three proteins with accession number BAA05496, CAA44268 and ANG90831 having II values greater than 40. The AI of the MBB proteins shows AAD09878, ANG90842 and ANG90831 having values greater than 100. The GRAVY values of the some MBB proteins some were -ve while otherswere +ve. The Ramanchandra plot shown in Figure 1 showed the protein favored glycine or is more of glycine which is an aliphatic amino acid.

Protein Accession	AA	Mwt	Pi	Q	EC	Halflife	П	AI	GRAVY
Rumber									
AAD09878	441	47611.32	9.53	+ve	43320	30 h	32.17	122.22	0.635
3I7J	280	29350.88	5.64	-ve	40450	Х	26.46	85.43	0.031
BAA05497	299	31088.89	6.13	-ve	83420	30 h	26.21	67.46	-0.049
BAA05496	338	35686.20	6.08	-ve	72880	30 h	40.28	73.08	-0.161
CAA44268	325	34546.84	5.62	-ve	81360	30 h	44.55	73.38	-0.176
ANG90842	408	42862.87	10.42	-ve	51910	30 h	26.97	124.34	0.753
ANG90839	495	55038.93	9.01	+ve	88350	30 h	31.73	80.38	-0.156
ANG90831	406	43815.26	9.02	+ve	30940	30 h	46.02	118.67	0.444
ANG90823	509	54268.68	5.29	-ve	42860	30 h	36.36	93.42	-0.019
ANG90818	718	80250.65	4.09	-ve	147140	30 h	32.81	82.47	-0.393

Table 1. Physico-chemical characteristic of MBB protein.

AA: Amino Acid; pI: Isoelectric Point; Q: Net Charge; II: Instability Index; AI: Aliphatic Index; GRAVY: Grand Average of Hydropathicity; EC: Extinction Coefficient; Mol wt: Molecular Weight



Figure 1. Ramchandran plot of selected Mycobacterium bovis protein sequence.

The amino acid composition of MBB proteins are shown in **Table 2**. All the proteins have high percentage composition of alanine, glycine and valine amino acids which belong to aliphatic amino acid group. All the proteins have zero percent composition in selenocystein and pyrrolysine which is the stop code.

Protein Accession Number	A	R	N	D	С	Q	E	G	н	Ι	L	К	М	F	Р	S	Т	W	Y	v	0	U
AAD0987 8	7.9	5.0	2.7	2.0	0.7	4.8	3.2	11.1	0.2	10.9	11.3	2.7	2.7	5.0	4.1	6.1	5.2	0.7	4.1	9.5	0.00	0.00
317J	13.9	5.0	1.8	5.4	0.7	2.5	4.3	11.1	4.6	1.8	9.6	1.1	0.0	3.9	6.1	5.4	7.9	2.1	1.8	9.3	0.00	0.00
BAA0549 7	16.7	3.3	5.7	4.3	0.0	3.3	1.7	14.4	2.3	2.6	6.7	1.7	4.3	4.0	6.4	6.7	3.7	4.3	2.7	5.4	0.00	0.00
BAA0549 6	10.9	3.6	4.4	5.3	0.9	5.3	2.1	13.6	1.2	1.8	8.9	3.3	3.0	3.6	6.8	7.4	4.4	3.0	3.6	7.1	0.00	0.00
CAA4426 8	12.0	3.4	5.5	4.3	0.9	4.9	2.8	12.3	0.9	3.4	8.9	2.8	3.1	3.1	6.8	8.9	3.7	3.4	4.3	4.6	0.00	0.00
ANG9084 2	14.0	5.9	2.2	2.7	0.0	2.5	1.5	10.3	2.0	5.6	14.5	1.0	3.9	2.9	4.9	4.9	6.4	1.7	2.2	11.0	0.00	0.00
ANG9083 9	9.5	8.9	2.2	6.1	1.8	2.6	3.4	8.7	3.6	4.2	7.5	2.2	2.6	4.4	6.1	5.1	6.9	2.4	3.0	8.7	0.00	0.00
ANG9083 1	12.8	9.4	0.2	4.4	1.0	3.2	4.9	7.9	2.0	5.4	13.3	1.0	2.5	3.2	3.4	5.9	5.7	1.0	1.5	11.3	0.00	0.00
ANG9082 3	13.8	7.3	2.9	5.5	2.0	3.5	6.3	9.6	1.8	4.5	9.6	2.0	1.6	2.6	5.5	4.1	5.5	0.8	2.8	8.4	0.00	0.00
ANG9081 8	9.7	7.7	2.9	8.4	0.4	2.5	6.5	8.2	2.6	3.6	8.9	2.5	1.1	2.9	5.7	4.87	5.8	2.4	5.0	8.2	0.00	0.00

 Table 2. Amino acid composition (%) of MBB proteins.

A: Alanine; R: Arginine; N: Asparagini; D: Aspartic Acid; C: Cysteine; E: Glutamic Acid; Q: Glutamin; G: Glycine; H: Histidine; I: Isoleucine; L: Leucine; K: Lysine; M: Methionine; F: Phenylalanine; P: Proline; S: Serine; T: Theonine; W: Tryptophan; Y: Tyrosine; V: Valine; U: Selenocystein; O: Pyrrolysine

The secondary structure prediction of MBB proteins is presented in **Table 3**. Proteins with the accession number AAD09878, ANG90842, ANG90831 and ANG90823 had high percent as alpha helice structure while the rest of the proteins had high percent in random coil structure.

Protein Accession Number	Alpha helice (%)	Extended Strand (%)	Beta Turn (%)	Random Coil (%)
AAD09878	34.08	31.29	12.02	22.68
3I7J	31.07	17.14	15.36	36.43
BAA05497	32.11	16.39	15.05	36.45
BAA05496	28.70	18.93	11.24	41.12
CAA44268	30.77	18.15	13.23	37.85
ANG90842	48.77	18.14	7.11	25.98
ANG90839	27.68	21.68	12.93	37.78
ANG90831	52.71	18.97	8.37	19.95
ANG90823	35.56	20.43	11.00	33.01
ANG90818	26.32	24.93	14.21	34.54

Table 3. Second	ary structure of MBB	proteins.
-----------------	----------------------	-----------

Parameters: Window Width: 17, Similarity Threshold: 8 and Number of States: 4

DISCUSSION

The computed isoelectric points (pI) for MBB will be useful for developing buffer system for purification by isoelectric focusing method. The isoelectric point is of significance in protein purification because it is the pH at which solubility is always minimal and at which mobility in an electro focusing system is zero and therefore the point at which the protein will accumulate [8]. The extinction coefficient of a protein at 280 nm depends almost exclusively on the number of aromatic residues, particularly tryptophan [9]. This indicates that the higher the EC value of the MBB proteins, the higher the number of aromatic residues [10,11]. In particular, hydrophobic amino acids can be involved in binding/recognition of hydrophobic ligands such as lipids [12]. All the MBB proteins have zero selenocystein and pyrrolysine which are interpreted as stop codons such that the protein cannot conclusively determine the identity of a residue [13].

Protein structure prediction from amino acid sequence is a fundamental scientific problem and it is regarded as a grand challenge in computational biology and chemistry. Given an amino acid sequence (i.e., the primary structure) which represents a monomeric globular protein in aqueous solution and at physiological temperatures, it is necessary to determine all helical segments and all beta-strands, all pairs of beta-strands which form beta-sheets (i.e., the beta-sheet topology), all disulfide bridges if cysteines are present, all loops that connect secondary structure elements, and the three-dimensional folded protein structure [14]. The accuracy of protein structure prediction depends critically on sequence similarity between the query and template as observed in the present study. If a template is detected with >30% sequence identity to the query, then usually most or all of the alignment will be accurate and the resulting relative positions of structural elements in the model will be reliable [5]. The practical applications of protein structure prediction are many and varied, including guiding the development of functional hypotheses about hypothetical proteins [15], improving phasing signals in crystallography [16], selecting sites for mutagenesis [17] and the rational design of drugs [18]. The present approach can be used to make functional predictions especially for newly discovered sequences [19].

CONCLUSION

The study finds out that the physico-chemical properties revealed the following; all the protein is high in EC, some of the proteins are high in II and AI. The amino acid compositions of all the protein are high in percent composition of alanine, glycine, leucine, isoleucine and valine amino acids. All the protein also showed high percent of alpha helice structure. Therefore, it may be concluded that *Mycobacterium bovis* is resistant to mutation and is thermally stable. I therefore recommend research on mutagenesis of this causative agent.

REFERENCES

- 1. Grange JM, Yates MD, De Kantor IN (1996) Guidelines for speciation within the *Mycobacterium tuberculosis* complex. 2nd Edn, WHO/EMC/ZOO/96.4.
- 2. Cole ST, Brosch R, Parkhill J (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393: 537.
- Dauda A, Duwa H, Yaska JA, Egahi JO, Kollo YS (2017) Protein sequence analysis of Newcastle (Paramyxovirus) in poultry. Int J Interdiscip Res Innov 5: 19-24.

- 4. Gasteiger E (2005) The Proteomics Protocols Handbook. Humana Press, pp: 571-607.
- 5. Kelley LA, Sternberg MJE (2009) Protein structure prediction on the web: A case study using the Phyre server. Nature Protocols 4: 363-371.
- Wass MN, Kelley LA, Sternberg MJ (2010) 3D ligand site: Predicting ligand-binding sites using similar structures. Nucleic Acids Res 38: 469-473.
- Leigh R, Ellis R, Wattie J, Southam DS, De Hoogh M, et al. (2002) Dysfunction and remodeling of the mouse airway persist after resolution of acute allergen-induced airway inflammation. Am J Resp Cell Mol Biol 27: 526-535.
- Fennema R (2008) Food Chemistry. 3rd Edn. CRC Press, pp: 327-328.
- 9. Gill SC, Von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 182: 319-326.
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, et al. (2003) ExPASy - The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31: 3784-3788.
- Munduganore DS, Mundaganore YD, Ashokan KV (2012) Sequence analysis of protein in pest des petits. Int J Food Agric Vet Sci.
- Betts MJ, Russell RB (2003) Amino acid properties and consequences of substitutions. In: Bioinformatics for Geneticists. John Wiley and Sons Ltd., pp: 289-316.
- 13. Suchanek M, Radzikowska A, Thiele C (2005) Photoleucine and photo-methionine allow identification of protein-protein interactions in living cells. Nat Methods 2: 261-267.
- 14. Floudas CA (2007) Computational methods in protein structure prediction. Biotechnol Bioeng 97: 207-213.
- Watson JD, Baker TA, Bell SP, Gann A, Levine M, et al. (2004) Molecular biology of the gene. 5th Edn. Peason Benjamin Cummings (Cold Spring Harbor Laboratory Press). ISBN: 08053-4635.
- Qian BD (2007) High-resolution structure prediction and the crystallographic phase problem. Nature 450: 259-264.
- 17. Rava P, Hussain MM (2007) Acquisition of triacylglycerol transfer activity by microsomal triglyceride transfer protein during evolution. Biochemistry 46: 12263-12274.
- Park H, Hwang KY, Oh KH, Kim YH, Lee JY, et al. (2008) Discovery of novel alpha-glucosidase inhibitors based on the virtual screening with the homology-

modeled protein structure. Bioorg Med Chem 16: 284-292.

 Ugbo SB, Yakubu A, Omeje JN, Musa IS, Bibinu BS, et al. (2015) Assessment of genetic relationship and application of computational algorithm to assess functionality of non-synonymous substitutions in DQA2 gene of cattle, sheep and goats. Open J Genet 5: 145-158.