ISSN 0189-0514
J. Anim. Prod. Res. (2017) 29(1):167-174

National Animal Production Research Institute
Ahmadu Bello University
P.M.B 1096, Shika-Zaria,
Kaduna State, Nigeria.
Email:japr@napri.gov.ng
Website: http://journals.napri.gov.ng

INSILICO ANALYSIS OF *MYCOBACTERIUM CAPRAE* PROTEIN SEQUENCE

[1]Dauda, A., [2]Abbaya, H. Y., [3]Abare, E. A. and [3]James, A. K
[1]Department of Animal Science University of Calabar P.M.B 1115 Calabar, Nigeria
[2]Department of Animal Production, Adamawa State University, P.M.B 25. Mubi, Nigeria.
[3]Department of Animal Science, University of Maiduguri P.M.B 1069, Maiduguri, Nigeria.
Corresponding author: Phone no: (+234)7036608339. Email: abbaya177@gmail.com

ABSTRACT

A total of ten (10) *Mycobacterium caprae* (MBC) proteins were retrieved from the GenBank (www.ncbi.nlm.nih.gov). The Genbank accession numbers of the sequences and sequence variations of the proteins were used to investigate the molecular identity of various MBC proteins. The physico-chemical properties of MBC proteins were performed using Protparam tool. 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 MBC proteins were acidic while some were basic in nature based on their pI. The EC of all the protein revealed appreciable value. The II of MBC protein value shows majority II<40. AI for some of the MBC protein have AI>100 while others AI<100. The GRAVY of MBC proteins revealed some are positive while some are negative. The amino acid composition of MBC proteins indicates that they are rich in aliphatic amino acids. The three dimensional structures (3D) of the MBC proteins were determine using Phyre2 server. The amino acid sequences of MBC proteins were subjected to secondary structure prediction using ExPASy's SOPMA tool. The MBC proteins some are alpha helice while others are random coil structure. The genetic information revealed in this study will help in carrying out research in mutagenesis and pharmacogenetic.

INTRODUCTION

Mycobacterium caprae is a species of bacteria in the genus *Mycobacterium* and a member of the Mycobacterium tuberculosis complex. Prior to 2003, the species was referred to as *Mycobacterium tuberculosis* subsp. Caprae (Rodríguez *et al.*, 2009). It is also synonymous with the name *Mycobacterium bovis* subsp. caprae. *M. caprae* is a causative agent of tuberculosis in humans (Kubica *et al.*, *2003*). The combination of disease tracing and molecular typing is needed to understand the epidemiology of tuberculosis. The aim of the study is to determine the sequence protein *Mycobacterium caprae* analysis.

Keywords: Protein, Mycobacterium caprae and Sequence

MATERIALS AND METHODS

A total of ten (10) *Mycobacterium caprae* (MBC) nucleotide sequences were retrieved from the GenBank (NCBI) (www.ncbi.nlm.nih.gov). The Genbank accession numbers of the sequences are AFZ62654, WP_ 054938511, WP_ 0549385671, WP_ 054938497, WP_ 054938541, WP_ 054938469, WP_ 054938610, WP_ 054938779, WP_ 054938588 and WP_ 054938520. ProtParam Tool was used for the computation of various physical and chemical properties of the MBC proteins using amino acid sequences. The computated parameters were molecular weight,

theoretical pI (isoelectric point), amino acid composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) (Gasteiger, 2005). The amino acid sequences of MBC proteins were subjected to secondary structure prediction using ExPASy's SOPMA tool. SOPMA 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 MBC proteins. These servers predict the three-dimensional structure of a protein sequence using the principles and techniques of homology modeling (Kelley and Sternberg, 2009). Currently, the most powerful and accurate methods for detecting and aligning remotely related sequences rely on profiles or Hidden Markov Models (HMMs). 3DligandSite was used to predict the binding site of the 3D structure of the MBC proteins. Phyre2 is coupled to the 3DligandSite server for protein binding site prediction (Wass *et al*., 2010).

RESULTS AND DISCUSSION

The variation in sequence length in base pair (bp) of MBC protein ranges between 432bp and 1232bp (Table 1). The variation in MBC protein sequence might results from evolution and differentiation (Yakubu *et al*., 2014). There are cases where variability might results from DNA duplication, DNA rearrangement, short tandem repeat (STR), insertions or deletion of sequences (Vincent *et al*., 2014). The variation might also be due to genome region of the protein molecules which enhances resistance to drug effect. New typing tool may help improve the surveillance and control of the disease, as well as to trace new epidemics.

Table 1: Accession Number and Sequence Length Variation of MBB protein

Protein Accession Number	Base Pair Number (Bp)	Sequence Length variation
AFZ62654	432	432 – 1232
WP_ 054938511	874	
WP_ 0549385671	1013	
WP_ 054938497	744	
WP_ 054938541	885	
WP_ 054938469	918	
WP_ 054938610	1110	
WP_ 054938779	861	
WP_ 054938588	1232	
WP_ 054938520	882	

The physical and chemical properties of MBC proteins are shown in Table 2. The amino acid number (AA) increase with increase in the molecular weight of the proteins (Mwt). The higher the amino acids number the higher the molecular weight. The isoelectric point (pI) - isoelectric point is the pH at which a protein carries no net charge. The isoelectric point is of significance in protein purification because it is the pH at which solubility is often minimal and at which

mobility in an electro focusing system is zero (and therefore the point at which the protein will accumulate. The pI of MBC protein with accession number AFZ62654 and WP_ 054938511 are basic with pI>7 while the rest are acidic pI<7. This measure indicates how much light is absorbed by a protein at a particular wavelength. The extinction coefficient of a protein at 280 nm depends almost exclusively on the number of aromatic residues, particularly tryptophan (Gill *et al*., 1989). This indicates that the higher the EC value of the MBC proteins, the higher the number of aromatic residues which made the protein highly stable (Gasteiger 2003). As a result of their stability, it is very difficult to cause aromatic molecules to break apart and to react with other substances (Hofmann, 1855). This is indicating resistance to mutation. The result of half-life of all the MBC proteins selected for the study is presented in Table 2. The result showed a value of 30 hours as the half-life of MBC protein. Half-life plays an important role in determining its stability in vivo (Bachmair *et al*., 1986). The instability index (II) of MBC protein is presented in Table 2. The result of the study revealed that protein with II<40 is stable while those with II<40 (Guruprasad *et al.,* 1990). The reason for the stability and the instability is dipeptides bond and the presence of such dipeptide bond make the significantly difference between the stable and unstable proteins. The aliphatic index (AI) of a protein is defined as the relative volume occupied by aliphatic side chains. Any protein with AI>100 is thermally stable while protein AI<100 is not thermally stable. (Munduganore *et al*., 2012). AI is regarded as a positive factor for the increase of thermal stability of globular proteins (Ikai, 1980). The Grand average hydropathicity (GRAVY) of MBC proteins showed some are positive while some are negative. The positive GRAVY value of the protein means that they are hydrophobic (not soluble in water) and have their surface rich with negatively charged amino acid like glutamate and aspartate amino acids while protein that have negative GRAVY value means they are hydrophilic (soluble in water) and have they surface rich with positively charged amino acids like lysine and arginine. All the proteins that are bind to positively charge DNA may regulate gene expression or fold the DNA (Urry, 2004). In particular, hydrophobic amino acids can be involved in binding/recognition (point where mutation have few or no effect) of hydrophobic ligands such as lipids (Betts *et al.*, 2003).

The prediction secondary structure of MBC protein is very important in determining the function and structure of the protein so as to enable easy classification of protein according to their structure. The prediction of secondary structure of MBC protein as shown in Table 3.The MBC protein selected for this study, some have high per cent of alpha helice and alpha helix preferred to be with aliphatic protein which save as a factor of increasing thermal stability while some of the protein are high in random coil percent. The practical applications of MBC protein structure prediction include guiding the development of functional hypotheses about hypothetical proteins, improving phasing signals in crystallography and selecting sites for mutagenesis (Qian *et al*., 2007; Rava and Hussain, 2007). The amino acids composition percentage is presented in Table 4. All the proteins have similar amino acids composition. The proteins have high per cent alanine, isoleucine and valine amino acid. These amino acids are aliphatic amino acid which is regarded as a factor of increasing thermal stability (Ikai, 1980). All the protein have zero per cent of Selenocystein and Pyrrolysine which is a stop code (identity of the protein cannot be further determined).

Table 2: Physico-chemical characteristic of MBC protein

Protein Accession Number	A A	Mwt	pI	Q	EC	Half-life	II	AI	GRAVY
AFZ62654	432	44293.38	10.74	+ve	70930	30	26.79	120.07	0.825
WP_ 054938511	874	96787.57	8.57	+ve	85830	30	41.96	89.11	-0.303
WP_ 0549385671	1013	80597.51	5.04	-ve	5500	30	9.93	31.34	-0.445
WP_ 054938497	744	75876.92	5.33	-ve	67380	30	39.95	75.08	-0.230
WP_ 054938541	885	72439.77	4.13	-ve	16960	30	14.17	59.38	0.208
WP_ 054938469	918	81290.27	4.28	-ve	31860	30	22.31	62.80	0.019
WP_ 054938610	1110	119430.19	5.52	-ve	82850	30	38.06	97.53	-0.110
WP_ 054938779	861	94254.77	4.91	-ve	130750	30	31.48	85.54	-0.153
WP_ 054938588	1232	136898.32	6.26	-ve	258520	30	39.09	72.81	-0.361
WP_ 054938520	882	95413.69	6.93	-ve	114820	30	46.62	86.01	-0.100

Key:
AA = amino acid
pI = isoelectric point
Q = net charge
II = instability index
AI = alphatic index
GRAVY = grand average of hydropathicity
EC = extinction coefficient
Mol wt = molecular weight

Table 4: The amino acid composition (%) of MBC protein

Protein Accession Number	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	O	U
AFZ62654	19.4	6.5	0.2	2.1	0.9	2.1	1.6	10.9	1.9	4.6	13.4	0.5	2.8	2.3	4.2	4.2	7.9	2.5	1.6	10.4	0.00	0.00
WP054938511	11.0	9.8	2.2	6.9	0.6	3.3	5.9	7.2	3.0	4.5	10.0	3.4	3.2	2.5	5.3	6.6	3.9	1.3	1.9	7.6	0.00	0.00
WP054938567 1	12.0	2.0	5.9	4.1	0.0	1.7	0.3	49.7	1.5	1.0	2.6	0.7	0.3	1.3	3.3	5.6	6.1	0.1	0.0	1.9	0.00	0.00
WP054938497	17.1	4.7	2.3	5.8	0.3	3.5	3.6	8.3	2.4	3.9	5.8	2.0	1.3	0.9	12.9	5.8	9.5	1.2	1.6	7.0	0.00	0.00
WP054938541	16.3	0.2	5.8	2.0	0.0	1.2	0.8	43.2	0.7	2.5	6.2	0.5	0.6	3.8	1.9	5.5	5.0	0.2	0.5	3.2	0.00	0.00
WP054938469	10.7	1.1	5.6	3.2	0.0	3.6	1.0	35.7	1.2	4.1	5.3	0.1	1.2	2.8	3.9	5.6	8.0	0.2	1.5	5.2	0.00	0.00
WP054938610	14.9	9.5	2.3	6.7	0.7	2.7	6.7	8.4	2.9	3.8	12.5	1.0	1.4	2.5	4.9	4.5	5.9	1.0	1.4	6.6	0.00	0.00
WP054938779	13.8	7.1	2.2	7.4	0.9	3.6	6.3	7.3	2.2	3.4	9.4	2.1	1.2	4.4	4.4	5.0	6.9	2.0	2.9	7.5	0.00	0.00
WP054938588	8.6	7.7	2.4	6.6	0.7	2.6	5.1	9.2	3.0	2.9	7.1	2.8	2.8	3.5	6.2	6.3	7.1	2.8	3.9	8.6	0.00	0.00
WP054938520	11.9	5.4	2.7	4.6	1.9	3.9	5.6	9.5	2.7	4.5	9.1	4.5	2.0	4.0	7.8	3.3	5.3	1.8	2.0	7.3	**0.00**	0.00

Key:
A = Alanine, Arginine = R, Asparagine = N, Aspartic acid = D, cysteine = C, Glutamic acid = E, Glutamin = Q, Glycine = G, Histidine = H, Isoleucine = I, Leucine = L, Lysine = K, Methionine = M, Phenylalanine = F, Proline = P, Serine = S, Theonine = T, Tryptophan = W, Tyrosine = Y, Valine = V, Selenocystein = U, Pyrrolysine = O.

Table 3: Secondary structure of MBC proteins

Protein Accession Number	Alpha helice (%)	Extended Strand (%)	Beta Turn (%)	Random Coil (%)
AFZ62654	48.15	22.22	7.87	21.76
WP_ 054938511	44.51	14.65	10.18	30.66
WP_ 0549385671	0.69	12.14	8.00	79.17
WP_ 054938497	37.77	9.14	6.50	46.51
WP_ 054938541	11.86	22.82	9.60	55.71
WP_ 054938469	9.37	24.18	11.22	55.23
WP_ 054938610	53.87	10.99	9.10	26.04
WP_ 054938779	44.60	16.96	10.57	27.87
WP_ 054938588	26.95	19.72	10.80	42.53
WP_ 054938520	33.79	13.83	8.96	43.42

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

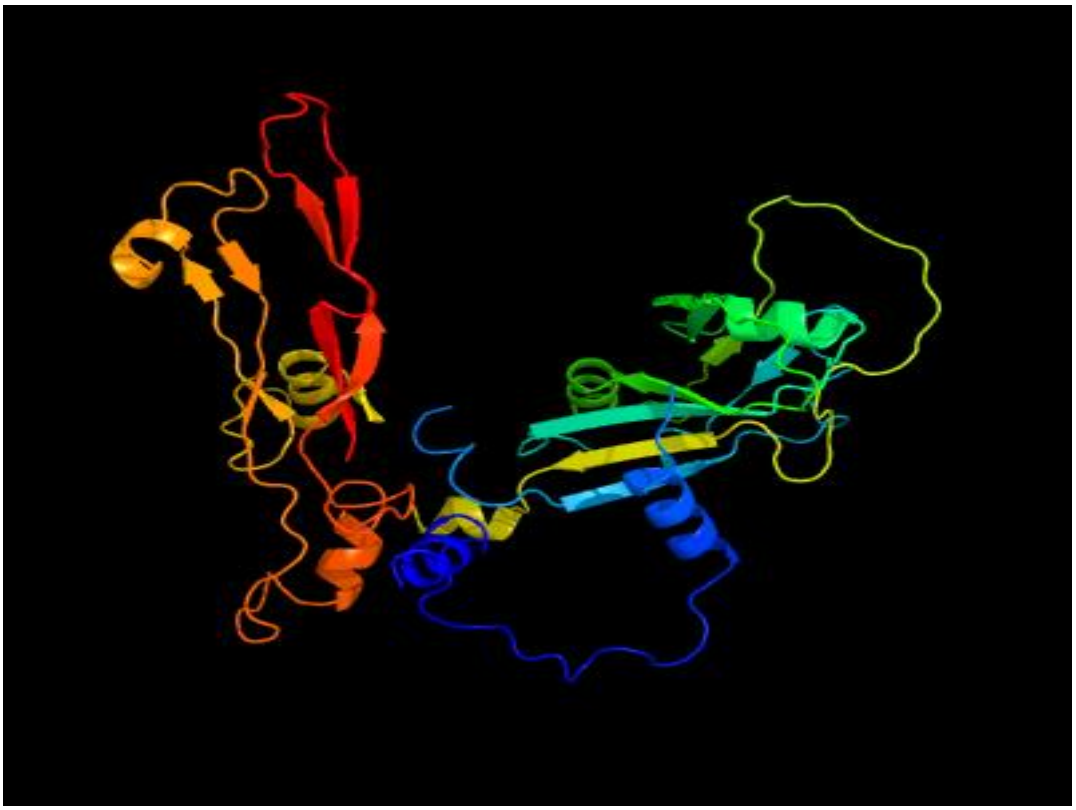


Figure 1. Schematic 3D structure of goat protein with accession number WP_ 054938520; Image coloured by rainbow N → C terminus; Model dimensions (Å): X: 68.312 Y:64.481 Z:61.432

CONCLUSION

The Genetic data revealed the sequence length variation, physical, chemical, amino acid composition and the secondary structure prediction of MBC protein which could bring new insights into epidemiological questions. 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. New typing tool may help improve the surveillance and control of the disease, as well as to trace new epidemics.

REFERENCES

Bachmair, A., Finley, D. and Varshavsky, A. 1986. In vivo half-life of a protein is a function of its amino-terminal residue. *Sciences,* 234: 179-186.

Betts, M.J. and Russell, R.B. 2003. Amino acid properties and consequences of substitutions. In *Bioinformatics for Geneticists* edition M.R. Barnes and I.C. Gray Chichester: Wiley.

Gasteiger, E. 2005. (In) John M Walker (edition). *The Proteomics Protocols Handbook.* Humana Press 571-607.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A. 2003. ExPASy - the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research.* 31: 3784-3788.

Gill, S.C. and Von Hippel, P. H. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182: 319–326.

Guruprasad, K., Reddy, B.V. and 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 England.* 2:155–61.

Hofmann, A. W. 1855. "On Insolinic Acid". *Proceedings of the Royal Society* 8: 1–3.

Ikai, A.J. 1980. Thermostability and aliphatic index of globular proteins. *Journal of Biochemistry* 88: 1895-1898.

Kelley, L.A. and Sternberg, M.J.E. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols,* 4 (3): 363–371

Kubica, T., Rusch-Gerdes, S. and Niemann, S. *2003. "*Mycobacterium bovis subsp. caprae caused one-third of human M. bovis-Associated tuberculosis cases reported in Germany between 1999 and 2001*". Journal of Clinical Microbiology. 41 (7): 3070–3077.*

Munduganore, D.S., Mundaganore, Y.D. and Ashokan, K.V. 2012. Sequence Analysis of Protein in Pest Des Petits. *International Journal of Food, Agriculture and Veterinary Sciences ISSN: 2277-209X*

Qian, B.D. 2007. High-resolution structure prediction and the crystallographic phase problem. *Nature,* 450: 259–264.

Rava, P. and Hussain, M.M. 2007. Acquisition of triacylglycerol transfer activity by microsomal triglyceride transfer protein during evolution. *Biochemistry*, 46: 12263–12274.

Rodríguez, E., Sánchez, L.P. and Pérez, S. 2009. "Human tuberculosis due to Mycobacterium bovis and M. caprae in Spain, 2004–2007". *International Journal of Tuberculosis and Lung Diseases. 13 (12): 1536–41*.

Urry, D. W. 2004. "The change in Gibbs free energy for hydrophobic association: Derivation and evaluation by means of inverse temperature transitions". *Chemical Physics Letters* 399 (1–3): 177–83.

Vincent, S.T., Momoh, O.M. and Yakubu, A. 2014. Bioinformatics analysis of beta-casein gene in some selected mammalian species. *Research Opinions in Animal and Veterinary Sciences*, 4(10): 564-570

Wass, M.N., Kelley, L.A. and Sternberg, M.J. 2010. 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Research*, 38: 469-473.

Yakubu, A., Alade, E.D. and Dim, N.I. 2014. Molecular analysis of Solute Carrier Family 11 A1 (SLC11A1) gene in ruminants and non-ruminants using computational method.*Genetika*, 46(3): 925-934.