International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 07
June 2017

Biocomputational Analysis of *Chlamydia abortus* Protein Sequence

Authors: [1]Ayuba Dauda, [2]Sinodo Saul and [3]Yerima Shettima Kolo

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

[2]Department of Animal Science University of Maiduguri P.M.B 1069 Maiduguri, Nigeria

[3]Department of Animal Science University of Agriculture P.M.B 2373 Makurdi, Nigeria

Corresponding email: ayubadauda87@gmail.com

Abstract

A total of fiften (15) protein sequence of *Chlamydia abortus* (*C abortus*) were retrieved from the GenBank (www.ncbi.nlm.nih.gov).The phyco-chemical properties of *C abortus* proteins were performed using protparam tool. The isoelectric point (pI), extinction coefficient (EC); instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY) were also computed. The study revealed that the pI value of *C abortus* protein showed that some were basic (>7) nature and acidic (<7) in nature respectively. The EC and II of protein showed that some *C abortus* protein have better stability which might be resistant to mutation. AI for all the protein showed that only *C abortus* protein with accession Number WP_072667807 showed AI > 100 which indicates thermally stable. The GRAVY of all the protein were negative (hydrophilic). The amino acid composition of *C abortus* proteins indicated high in serine and threonine which are hydroxyl amino acid which is non reactive and can play a role in substrate recognition.. The prediction of secondary structure was performed using SOPMA. The proteins are more of random coil structure then followed by alpha helix. Phyre2 server was used to predict the 3D structure of *C abortus* proteins. Molecular analysis should be carry out to substantiate this findings.

Keywords: Protein, *Chlamydia abortus* and Mammals

Introduction

Chlamydophila abortus is an intracellular Gram negative bacterial pathogen that is endemic throughout the world. *C. abortus* is the most common cause of infectious abortion in mammals (sheep, cattle, pig and goats) in farm animals. It also represents a significant zoonotic risk to pregnant women. The diseases caused by both pathogens result in enormous economic costs to their respective livestock industries. *Chlamydophila abortus* is a member of the *Chlamydiaceae*, a phylogenetically distinct Gram negative bacterial family, encompassing two genera (*Chlamydia* and *Chlamydophila*), which are subdivided into three (*Chlamydia muridarum*, *Chlamydia suis*, and *Chlamydia trachomatis*) (Everett *et al*., 1999). In Nigeria with the outbreak of this deadly

disease had cause a serious loss in poultry industry. Although much veterinary work had been done on this disease but little or no effort has been made genetically in Nigeria. The aim of this study is to carry out protein sequence analysis in order to provide genetic information which may help to curtail the effect of this disease in farm animals.

Materials and Methods

A total of fifteen (15) protein sequence of *Chlamydia abortus* (CA) were retrieved from the GenBank (www.ncbi.nlm.nih.gov). The Genbank accession numbers of the sequences and amino acid based pair number are presented in Table 1.

Table 1. Protein Accession Number and Amino Acid Based Pair Number

S/N	Chlamydia abortus Protein Accession Number	Amino Acid Based Pair Number
1	WP_072667782	1533
2	WP_006343931	1024
3	WP_006343930	980
4	WP_0066343927	988
5	WP_072667881	700
6	WP_072667879	1378
7	WP_072667861	732
8	WP_072667858	908
9	WP_072667842	790
10	WP_072667837	806
11	WP_072667827	1139
12	WP_072667807	864
13	WP_072667796	1393
14	WP_072667787	1045
15	WP_072667780	1005

ProtParam Tool was used for the computation of various physical and chemical properties of the *Chlamydia abortus* proteins using amino acid sequences. The computated parameters were molecular weight (MW), theoretical (pI) (isoelectric point), extinction coefficient (EC), estimated half-life (EHL), instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY) (Gasteiger, 2005). The amino acid sequences of *Chlamydia abortus* protein 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). It predicts the secondary structure by consensus prediction from multiple alignments. SVMProt tool will be used to predict the functional signature of the selected *Chlamydia abortus* protein sequences. SVMProt classifies a particular protein into its functional family from its primary sequence (Cai *et al.*, 2003.The Phyre2 server was used to predict the 3D structure of *Chlamydia abortus* 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 *Chlamydia abortus* proteins. Phyre2 is coupled to the 3DligandSite server for protein binding site prediction (Wass *et al*., 2010).

Results and Discussion

The result of physical and chemical parameters of *Chlamydia abortus* were presented in Table 2. The molecular weight of the protein increase with the in amino acid number. the isoelectric point (pI) is the pH of the *Chlamydia abortus* protein. The pI value < 7 is acid and value > 7 is basic. The isoelectirc point (pI) is the 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 (Fennema, 2008). The net charge (Q) of the *Chlamydia abortus* protein showed positive, negative and neutral. The net charge of the protein that showed positive charge means they are extracellular protein and the protein showed negative charge means intracellular protein (Munduganore *et al*., 2012). The extinction coefficient (EC) 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 protein, the higher the number of aromatic residues which made the protein highly stable (Gasteiger 2003; Munduganore *et al*., 2012). This implies that this causative organism that cause abortion in mammals is resistant to mutation therefore, a post modification translation is require to alter the chemical composition of the host organism or the therapy/vaccine of the causative agent. The half life of all the *Chlamydia abortus* protein were 30 hours. the half life is the time taken of the amount of protein in a cell to disappear after its synthesis in the cell. The instability index (II) provides an estimate of the stability of protein in a test tube. II value below 40 is predicted as stable, a value above 40 predicts that the protein may be unstable (Guruprasad *et al.*, 1990). The proteins that have II value below 40, implies they are stable and resistant to mutation from generation to generation while the proteins with II value above 40 are not stable they have been mutated. The aliphatic index (AI) of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It may be regarded as a positive factor for the increase of thermostability of globular proteins (Ikai, 1980). AI above 100 indicates thermo stability while values below 100 are having less thermo stability. in this study only Chlamydia

abortus protein with accession number WP_072667807 is having AI value above 100. This indicated that about 99% of proteins selected for this study are thermally unstable. This implies that use of heart will help in curtailing the effect of the organism. The average grand hyropathicity (GRAVY) in this study showed negative value which means the protein selected for this study are all hydrophobilic which mean they are soluble in water.

Table 2: Physical and chemical characteristics of protein of Chlamydia abortus Protein

Protein Accession Number	AA	MolWt	pI	Q	EC	Half Life	II	AI	GRAVY
WP_072667782	1533	163571.17	5.27	-ve	118610	30hrs	30.31	81.37	-0.208
WP_006343931	1024	108777.57	8.20	+ve	96720	30hrs	29.79	79.67	-0.155
WP_006343930	980	104958.62	8.40	+ve	104170	30hrs	31.33	77.15	-0.181
WP_0066343927	988	108371.87	6.83	+ve	126060	30hrs	43.32	80.40	-0.246
WP_072667881	700	74234.13	5.28	-ve	91220	30hrs	29.36	70.40	-0.315
WP_072667879	1378	144978.14	5.27	-ve	42770	30hrs	42.56	70.99	-0.396
WP_072667861	732	80217.21	5.41	-ve	42860	30hrs	43.82	89.74	-0.152
WP_072667858	908	103483.81	5.33	-ve	120560	30hrs	50.17	87.06	-0.281
WP_072667842	790	89254.87	8.30	+ve	89620	30hrs	29.72	87.08	-0.289
WP_072667837	806	90471.21	5.58	-ve	60170	30hrs	40.40	91.45	-0.298
WP_072667827	1139	127351.79	7.34	Neutral	111160	30hrs	46.15	97.24	-0.068
WP_072667807	864	96971.79	5.40	-ve	60280	30hrs	41.72	104.42	-0.230
WP_072667796	1393	155010.23	7.53	+ve	90650	30hrs	36.92	99.55	-0.239
WP_072667787	1045	119099.32	5.41	-ve	169030	30hrs	39.32	87.36	-0.370
WP_072667780	1005	113855.04	5.84	-ve	137630	30hrs	39.33	95.60	-0.206

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

The amino acid composition of *Chlamydia abortus* are presented in Table 3. All the Chlamydia abortus proteins showed high values of amino acid composition in glycine, leucine, alanine, threonine and serine. Glycine, leucine and alanine are aliphatic amino acid which are non reactive and rarely involved directly in protein function, though they can play a role in substrate recognition (Barnes *et al*., 1999), a point where mutation have less effect. The hydroxyl group is fairly reactive, and can form hydrogen bonds with a variety of polar substrates (Barnes and Russell, 1999). Selenocystein and Pyrrolysine is zero for all the protein which a stop code (identity of the cannot be further determine). The secondary structure prediction of *Chlamydia abortus* protein is presented in Table 4. All the proteins are high in percent composition of random coil than followed by alipa helice structure. The practical applications of protein structure prediction are many and vary, include guiding the development of

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 07
June 2017

functional hypotheses about hypothetical proteins, improving phasing signals in crystallography and selecting sites for mutagenesis (Qian *et al*., 2007; Rava and Hussain, 2007).

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 07
June 2017

Table 3. Amino Acid Composition of Chlamydia abortus 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
WP_072667782	7.7	2.7	6.7	4.8	1.2	3.3	6.1	10.5	2.5	5.5	8.3	5.0	1.1	4.8	2.9	11.0	5.7	0.7	2.5	6.7	0.0	0.0
WP_006343931	8.7	3.1	8.9	4.1	1.5	2.7	3.0	9.0	1.8	6.2	7.6	4.4	1.0	4.8	3.9	11.6	8.1	1.0	2.7	6.0	0.0	0.0
WP_006343930	7.7	2.8	8.0	3.1	1.6	2.8	3.7	8.9	2.3	5.9	7.7	4.6	0.6	5.1	5.1	10.7	9.5	1.0	3.4	5.7	0.0	0.0
WP_0066343927	6.9	3.3	8.3	3.6	2.1	2.9	4.5	7.0	1.9	7.2	8.8	4.6	1.1	4.5	4.8	10.6	8.5	1.1	4.5	3.8	0.0	0.0
WP_072667881	7.0	1.3	7.4	4.9	1.9	2.9	4.1	10.6	1.4	5.6	7.3	5.7	0.7	4.0	4.0	11.7	9.7	1.3	4.0	4.6	0.0	0.0
WP_072667879	9.9	1.7	6.2	4.4	0.9	3.8	6.4	7.3	2.2	5.7	5.6	6.1	0.8	3.6	5.7	12.6	8.8	0.5	1.7	6.0	0.0	0.0
WP_072667861	7.0	5.6	4.5	5.6	2.3	2.9	7.2	6.7	1.9	4.9	10.1	4.8	2.7	3.4	3.1	9.7	6.7	0.5	1.9	8.3	0.0	0.0
WP_072667858	5.5	4.4	3.0	5.3	2.0	5.6	7.9	5.9	2.9	5.3	10.8	5.3	2.1	4.8	5.0	7.8	4.0	1.1	4.8	6.5	0.0	0.0
WP_072667842	3.8	4.2	5.4	5.2	0.9	2.8	6.2	8.4	2.0	6.3	10.5	7.7	0.9	6.3	4.3	6.7	6.7	0.8	4.8	6.1	0.0	0.0
WP_072667837	6.5	5.2	4.2	6.1	1.2	3.7	8.3	7.6	1.7	7.8	9.2	7.2	2.1	4.1	3.0	6.3	5.0	0.2	4.1	6.5	0.0	0.0
WP_072667827	5.1	2.8	3.6	3.9	0.9	4.4	6.3	5.6	2.8	8.2	11.7	7.4	0.5	5.9	5.3	11.0	5.4	1.0	3.0	5.1	0.0	0.0
WP_072667807	8.0	5.9	3.1	5.9	0.6	3.0	10.2	5.7	1.2	7.1	12.8	8.0	2.0	3.1	3.8	6.8	3.2	0.6	2.5	6.5	0.0	0.0
WP_072667796	5.8	5.7	3.0	5.9	1.4	3.2	8.1	8.4	2.2	8.3	10.1	8.4	1.9	2.8	3.9	5.2	5.0	0.5	2.5	7.6	0.0	0.0
WP_072667787	7.1	5.0	5.0	6.5	1.4	3.7	7.8	5.2	1.7	7.5	8.7	6.9	2.4	3.3	2.9	7.0	5.8	1.7	4.5	5.9	0.0	0.0
WP_072667780	7.3	4.5	3.9	5.6	1.1	4.5	6.7	4.8	3.2	6.1	12.4	5.6	1.5	4.4	4.5	7.3	6.2	1.5	3.7	5.6	0.0	0.0

A=Alanine, Arginine=R, Asparagini=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 4. Prediction of secondary structure of Chlamydia abortus protein

Protein Accession Number	Alpha Helice (%)	Extended Strand (%)	Beta Turn (%)	Random Coil (%)
WP_072667782	19.96	28.51	11.48	40.05
WP_006343931	16.87	31.98	9.90	39.16
WP_006343930	17.24	28.88	11.12	42.76
WP_0066343927	19.94	28.54	12.25	39.27
WP_072667881	12.71	34.29	12.00	41.00
WP_072667879	18.29	24.02	8.27	49.42
WP_072667861	46.58	13.93	10.93	28.55
WP_072667858	41.41	17.51	10.46	30.62
WP_072667842	24.56	29.49	10.89	35.06
WP_072667837	39.21	20.84	12.41	27.54
WP_072667827	27.83	28.09	9.75	34.33
WP_072667807	57.99	14.24	7.87	19.91
WP_072667796	36.40	22.54	11.13	29.94
WP_072667787	46.22	19.04	9.95	24.78
WP_072667780	46.87	15.02	8.06	30.05

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

Figure 1. showed the 3 dimensional structure of *Chlamydia abortus* protein. The 3 dimensional structure is the recent progress in predicting the full 3-D structure of transmembrane proteins (Yarov-Yarovoy *et al*., 2006), the most widely applied prediction technique for these proteins is to determine the transmembrane topology, i.e. the inside–outside location of the N and C termini relative to the cytoplasm, along with the number and sequence locations of the membrane spanning regions. This will facilitate the understanding of the structure and function of *Chlamydi abortus* proteins.

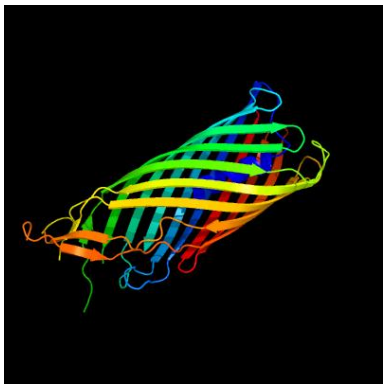

Image coloured by rainbow N → C terminus
Model dimensions (Å): **X**:44.707 **Y**:38.518 **Z**:71.590
Figure 1. 3D structure of *Chlamydia abortus* Protein

Conclusion

The study revealed genetic information about the physical and chemical properties; amino acid composition, secondary structure and the 3 dimensional structure on the causative agent of abortion in mammals

especially farm animals. This Genetic data may bring new insights into epidemiological questions. Molecular typing which will be instrumental in determining the population structure and evolution of pathogens. Since abortion in farm animals has economical consequences, efforts should be intensified towards finding sustainable genomic solutions to this disease 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. Molecular research should be carry out to substantiate this finding.

Reference

Barnes, M. R. and Russell, R. B. (1999). A lipid-binding domain in Wnt: a case of mistaken identity? *Curr Biol* 9(19):717–719.

Cai, C.Z., Han, L.Y., Ji, Z.L., Chen, X. and Chen, Y.Z. (2003). SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucl. Acids Res*., 31 (13): 3692-3697.

Everett, K. D. E., Bush, R. M. and Andersen, A. A. (1999). Emended description of the order Chlamydiales, proposal of Parachlamydiaceae fam. nov. and Simkaniaceae fam. nov., each containing one monotypic genus, revised taxonomy of the family Chlamydiaceae with description of five new species, and standards for the identification of organisms. Int J Syst Bacteriol 49, 415–440

Fennema. (2008). *Food Chemistry 3rd Ed*. CRC Press. pp. 327–8.

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. *NucleicAcids Res.* 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 Eng.* 2:155–61.

Ikai, A.J. (1980).Thermostability and aliphatic index of globular proteins. *J Biochem* 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.

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., Raman, S., Das, R., Bradley, P., McCoy, A.J., Read, R.J. and Baker, 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.

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.

Yarov-Yarovoy, V., Schonbrun, J. and Baker, D. (2006). Multipass membrane protein structure prediction using Rosetta. Proteins, 62: 1010–1025.