COMPARATIVE GENOMICS OF GENE S100A8 OF PIG AND ITS PROTEIN PRESCRIPTION WITH OTHER MAMMALS

O. O. OLUWOLE

Institute of Agricultural Research and Training, Obafemi Awolowo University, Nigeria.

[1]*Correspondence:* oluwafunmike@yahoo.co.uk +234(0)7031567164

ABSTRACT

Gene S100A8 is the gene that encodes the Protein S100. It is a family of low-molecular-weight protein found in vertebrates and characterized by two calcium-binding sites that have helix-loop-helix conformation. It plays a prominent role in the regulation of inflammatory processes and immune response. The nucleotide and amino acid sequence (AAS) of Gene S100A8 of pig were downloaded from the National Center for Biotechnology information (NCBI) data base, United State of America and Uniprot. The Gene S100 AAS of other species were downloaded from Universal protein resources (Uniprot) database, United Kingdom. Data generated were analyzed by bioinformatics tools. The result shows that the protein structure of 87 residues (98%) modelled at >90% accuracy was obtained. The secondary structure disorder prescription result was 22% disorder, Alpha Helix of 70% and 0% of Beta strand. The evolution relationship of Gene S100A8 of pig and other mammalswere highly related and conserved among the mammalian species; they have high comparability and they evolved from common ancestors.

Keywords: Pig, Gene S100A8, Protein S100, protein structure, phylogenetic tree.

INTRODUCTION

S100A8 is the gene that encodes Protein S100 (*Schäfer and Heizmann (1996).* Protein S100 is a small calcium and zinc binding protein. It is a Calcium binding protein that participates in calcium cell signaling pathways by binding to Ca^{2+}. This Ca^{2+} plays an important role in many cellular processes. It has unique specific domains that bind to Calcium binding proteins (CBP) and is known to be heterogenous (*Schäfer and Heizmann, 1996).*

S100 protein is a family of low-molecular-weight protein found in vertebrates and characterized by two calcium-binding sites that have helix-loop-helix conformation. There are at least 21 different S100 proteins (S100A1, A2…etc). Some of them are homodimeric, consisting of two identical polypeptides, which are held together by non-covalent bonds. It plays a prominent role in the regulation of inflammatory processes and immune response. Its pro-inflammatory activity involves recruitment of leukocytes, promotion of cytokine and chemokine production, and regulation of leukocyte adhesion and migration (http://www.uniprot.org/uniprot/C3S7K5). It also acts as a monocyte and mast cell chemo-attractant, it can stimulate mast cell degranulation and activation which generates chemokines, histamine and cytokines inducing further leukocyte recruitment to the sites of inflammation. This protein has been discovered in fish recently and has been emphasized as the potential marker for mineralizing cartilage in developing fish (Fonseca *et al.,* 2011).

This study compares the Gene S100A8 of pig with Gene S100 of other species and to predict its structure.

MATERIALS AND METHODS

The nucleotide and amino acid sequence (AAS) of Protein S100 of pig were downloaded from the National Center for Biotechnology information (NCBI) data base, United State of America and Uniprot. The Protein S100 AAS of other species were downloaded from Universal protein resources (Uniprot) database United Kingdom. The Identity and Similarity percentage of AAS of Protein S100 of pig and Protein S100 of other mammals were identified by conducting pairwise comparison of their AAS using two or more sequences of Basic local alignment search tool (BLAST).The evolutionary history was inferred using the Neighbor-Joining method (Saitout*et al.*, 1987). The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and were in the units of the number of amino acid substitutions per site. The analysis involved 29 amino acid sequences. All positions containing gaps and missing data were eliminated.Evolutionary analyses were conducted in MEGA7 (Kumar *et al*, 2016)

Results and Discussions

The sequence variations observed in camel, Chinese tree shrew, Rhesus Macaque, Green Monkey, chimpanzee, Gorilla, Pongo Abelii, Naked mole rat, Draland mole rat and that of Gene S100 of pig were probably by deletion or insertion of some AA sequences due to convergent evolution.

The protein structure of 87 residues (98%) modelled at >90% accuracy was obtained. The result can be viewedhttp://www.sbg.bio.imperial.ac.uk/3dligandsite/3dligand_report_p2.cgi. The secondary structure disorder prescription result was 22% disorder, Alpha Helix of 70% and 0% of Beta strand. The first template gave 99% confidence. The result showed that it is an EF hand fold and super family of EF hand that prove its true identity of Protein S100.

The protein parameters of mammals with different AAS length compared with that of pig AAS length of 89 were obtained from Expasy Bioinformatics resource portals where Protfam tool was used to obtain the protein parameters as shown in Table 2. The results obtained show that mammals with instability index greater than 40 caused the protein to be unstable.

The gene identity and similarity between the AAS of Gene S100A8 of Pig and Gene S100A8 of other mammals is shown in Table 3. The result shows that Gene S100A8 of pig shared percent identity ranging from 61.4% to 85% at AAS levels with other species Gene S100A8 sequences. The lower percentage observed between the Gene S100A8 of pig and other species such as David myotis, maydisbrandt bat, Chinese hamper, Guinea pig, Naked mole rat and Draland mole rat (61.4 to 63.6) revealed that they are not performing te same function (Durosaro et al., 2016; Baxevanis and Quellete, 2001). Gene S100A8 of other animals has higher percentages similarities with Protein S100 of pig ranging from 79.5% to 96.6%. The Giant panda and cattle had the highest similarity percentage while Draland mole rat had the lowest percentage

The higher Similarity percentage obtained between the GeneS100A8 of pig and other species implied that they all have common ancestors and homologous since the expected-value (e-value) were low for all the species, therefore they were significant (less than 0.05) and homologous (Lesk, 2002).

The phylogenetic tree obtained from the relationship between Gene S100A8 of pig and others animals was shown in Figure 2.(Kumar *et al.*, 2016).There were a total of 88 positions in the final dataset. The optimal tree with the sum of branch length of 2.89 was shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) was shown next to the branches (Felsenstein*et al,* 1985).

The phylogenetic tree derived from this study show that Gene S100A8 of pig and other species shared five common ancestors. The dendrogram obtained revealed that Gene S100A8 of different species was clearly separated from each other based on their different taxonomic classes.

In conclusion, this study shows that Gene S100A8 was highly related and conserved among the mammalian species, had high comparability and evolved from common ancestors. This study gives basic information that may be useful in the development of diseases resistance because of its pro-inflammatory activity that involves recruitment of leukocytes, promotion of cytokine and chemokine production, and regulation of leukocyte adhesion and migration that is useful in the regulation of inflammatory processes and immune response. It can also be used as biomarker for diseases detection in the future.

REFERENCES

Amills, M. (2010).Biodiversity and Origin of Pig Breeds.Bulletin of the University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca. Animal Science and Biotechnologies 68 (1/2) 1-5.

Amills, M., Ramires, O., Omitogun, O.G and Clop, A. (2012).Domestic pigs in African Africa Acheological Review.*Springer*, doi10.10007s10437,-012-9111-2. 10pp.

Baxevanis A.D and Quellete B.F.F. (2001).Bioinformatics: A practical guide to

Analysis of gene and protein. 2nd edition John Wiley and sons, Inc USA. 470pp.

Durosaro, S.O., Ilori, B.M, Sanda, A.J Onagbesan O.M, Peters S.O and Ozoje M.O (2016). Mutational and evolutionary Analysis of Bovine Reprimo Gene. Nigerian Journalof animal production.The Nigerian society for animal production. Vol 43(1), page 10-19

Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.

Fonseca VG , Fonseca VG , Rosa J , Laizé V , Gavaia PJ , Cancela ML (2011). Identification of a new cartilage-specific S100-like protein up-regulated during endo/perichondral mineralization in gilthead seabream.Gene Expression Patterns : GEP [2011, 11(7):448-455]

Josh T. and Xu D. (2007).Quantitative assessment of relationship between sequence similarity and function similarity.Biomed Central Genomics. 8;222-234.

Kumar S., Stecher G., and Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets.Molecular Biology and Evolution 33:1870-1874.

Lehninger A.L., Nelson, L.D. and Cox, M M. 1993.Principe of Biochemistry 2nd edition.Worth publisher Inc. USA. 432pp

Lesk, A.M (2002). Introduction to Bioinformatics 2nd Edition.Oxford University press, Oxford, 255pp.

Nacken W, Roth J, Sorg C, Kerkhoff C (2003). "S100A9/S100A8: Myeloid representatives of the S100 protein family as prominent players in innate immunity.".Microsc. Res. Tech. **60** (6): 569–80. doi:10.1002/jemt.10299

Saitou N. and Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425.

Schäfer BW, Heizmann CW (1996). "The S100 family of EF-hand calcium-binding proteins: functions and pathology.".Trends Biochem. Sci. 21 (4): 134–40.

Uniprot 2017.Protein database. http://www.uniprot.org/uniprot/C3S7K5

Zuckerkandl E. and Pauling L. (1965).Evolutionary divergence and convergence in proteins.Edited in Evolving Genes and Proteins by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York.

Table 1: The percent identity and similarity between the AAS of S100 protein of Pig and S100 protein of other mammals.

Animals	Accession Number	Class of Animal	AASL	Similarities (%)	Identity (%)	e-value
Pig	C3S7K5	Mammalian	89	100	100%	1.7e-57
Giant Panda	D2HJR9	Mammalian	89	96.6	85.4	3.8e-51
Cattle	P28782	Mammalian	89	96.6	83.44	.4e-50
America mink	U6CQ89	Mammalian	89	95.5	82	3.6e-49
Wild Yak	L8HQU7	Mammalian	89	95.5	82	3.6e-49
Dog	C0LQL0	Mammalian	89	95.5	83.1	3.6e-49
MustelaFuro	M3YAE9	Mammalian	89	95.5	80.9	4.6e-48
Black flying fox	L5JX40	Mammalian	89	93.3	79.8	4.6e-48
Buffalo	E1AHZ7	Mammalian	89	93.3	79.8	4.6e-48
Horse	F6YYI3	Mammalian	89	94.4	80.9	4.6e-48
Marmoset	U3CKF6	Mammalian	89	88.8	79.8	2.1e-45
Camel	S9XE92	Mammalian	93	93.3	77.5	2.4e-45
Rabbit	G1SXG6	Mammalian	89	90.9	75	3e-45
Otoga	H02H36	Mammalian	89	88.8	77.5	4.9e-44
Green Monkey	A0A0D9S5P	Mammalian	93	91.0	74.2	6e-43
Rhesus Macaque	F6QJD8	Mammalian	93	89.9	75.3	1.6e-43
Chinese tree shrew	L9JAM5	Mammalian	135	89.9	71.9	4.2e-42
Pongo Abelii	H2N5P6	Mammalian	93	88.8	74.2	2.5e-42
Gibbon	G1RHF0	Mammalian	93	89.9	71.9	3.6e-42
Gorilla	G3RWR9	Mammalian	93	89.9	73	3.6e-42
Chimpazee	H2Q028	Mammalian	93	89.9	71.9	1e-41
Human	P05109	Mammalian	93	89.9	71.9	1e-41
Squirrel	I3MT07	Mammalian	89	86.5	73	2.6e-41
David mayotis	L5MCF4	Mammalian	89	88.8	62.9	1.5e-37
Maydisbrandt bat	S7MV67	Mammalian	89	88.8	61.8	2.1e-37
Guinea pig	H0UWS1	Mammalian	91	87.5	61.4	1.8e-36
Chinese hamper	A0A061IM07	Mammalian	89	87.6	61.8	2e-35
Naked mole rat	G5BIG3	Mammalian	89	80.7	63.6	1.6e-34
Draland mole rat	A0A091DPX0	Mammalian	92	79.5	62.5	1.4e-33

AAS =Amino acid sequence

Table 2: Protein parameters of pig and other mammals

Protein Parameters	Pig	Camel	Human	Chinese tree shrew	Guinea pig	Draland Mole rat
AAS length	89	93	93	135	92	91
Carbon C	473	492	490	684	477	482
Hydrogen H	736	774	764	1072	749	757
Nitrogen N	116	128	128	180	133	125
Oxygen O	142	141	204	204	139	139
Sulfur S	3	3	2	9	3	3
Total atoms	1470	1538	1530	2149	1501	1506
Instability index:	42.61	25.04	21.37	46.51	36.56	31.52
Stability	Unstable	Stable	Stable	Unstable	Stable	Stable
Aliphatic index:	97.53	96.45	101.72	85.85	83.70	96.48
Hydropathicity (GRAVY)	-0.397	-0.397	-0.462	-0.283	-0.592	-0.408

AAS =Amino acid sequence, c=carbon, H =hydrogen, N=nitrogen, O=oxygen, Sulphur.

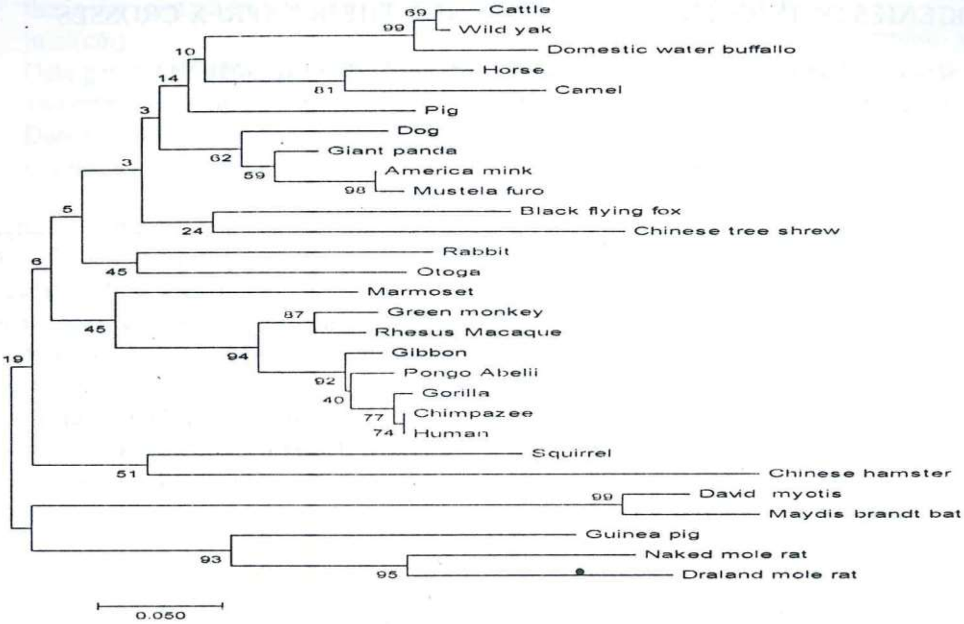

Figure 2: Phylogenetic tree of Protein S100 of pig with other species'