IDENTIFICATION OF FUNCTIONAL DOMAINS IN PROLACTIN GENE AND ANNOTATION OF MARKERS FOR THEIR FUNCTIONAL RELEVANCE

Sudhir Verma* and P Agastian

Research and Development Centre,
Bharathiar University, Coimbatore – 641 046
Loyola College, Dept of plant biology and Biotechnology, Chennai - 600 034
*Corresponding Author: sudhirpbt01@yahoo.com, sudhirverma22@hotmail.com

ABSTRACT

Introduction
Studying functional domains/markers is of relevance in understanding the domains which are indispensible for the gene's milk producing capacity or other milk quality associated attributes. This study involves identification of functional domains in the prolactin gene in four mammals namely cow (Bos taurus), chimpanzee (Pan troglodytes), sheep (Ovis Aries) and human (Homo sapiens).

Materials & Methods
Sequencing data generated from lab also compared and aligned with NCBI data. The same amino acid sequence generated from NCBI ws used for identification of functionl domain in protein sequence. Prolactin amino acid sequences of chimpanzees, sheep, cow and human were retrieved from genbank (NCBI). These were used to construct protein models using Swift Modeller software. The sequences and structures were aligned using Visual Molecular Dynamics (VMD) software. The protein models were analysed. The sequence and structure based alignments were used to determine signature domains/functional domains/markers in the prolactin gene using 3D structure modelling and ExPASy tool. We also performed structural and sequence based phylogenetic analysis.

Results
The structural and sequence based analysis confirmed the presence of two functional motifs in the prolactin gene viz. 'CHTSSLPTPEDKEQAQQT

HHEVLMSLILGLLRSW' and 'CLRRDSSKIDTY

LKLLNC'. The prolactin protein in cow, sheep and chimpanzees exhibit a similar protein structure which is evident from the RMSD and Qres values. Human prolactin displays structural differences when compared with other mammals.

Conclusion
The human prolactin protein displays slight structural differences when compared with prolactin of sheep, cow and chimpanzees, but interestingly, the functional motifs remain conserved.

Keywords: Sequencing Analysis, Multiple Sequence Alignment, Prolactin

INTRODUCTION

Studying functional domains/markers is of relevance in understanding the domains which are indispensible for the gene's milk producing capacity or other milk quality associated attributes. Prolactin (PRL) is a significant contributor to major milk components including

milk proteins, lactose and lipids. It is also involved in development, differentiation and regulation of the mammary gland.

Previous studies have proposed several genes as potential candidates for dairy traits, and *prolactin* gene (PRL) seems to be the most promising since it plays a crucial role in mammary gland development, initiation and maintenance of lactation, and expression of milk protein genes. Prolactin stimulates mammary development and promotes the formation and action of the corpus luteum during the female reproductive cycle in mammals. Allelic variations in the structural or regulatory sequences of PRL and variations in genes upstream and downstream to PRL in lactation pathway are of interest because of their possible direct or indirect effect on milk production [1-4].

Holeisten Fries, the cattle breed known for its high milk yield, average fat percentage, and low SNF percentage, raised in the Northern part of India exhibited highest expression of the prolactin gene, and is known for its high milk yield [5].

Structure is primarily responsible for the functional features displayed by a protein. The conservation of tertiary structure clearly implies the functional significance of a motif. The functional motif remains conserved during the process of evolution, and experience a selection pressure, which preserves the functional motif. This study involves identification of functional domains in the *prolactin* gene in four mammals namely cow (*Bos taurus*), chimpanzee (*Pan troglodytes*), sheep (*Ovis Aries*) and human (Homo sapiens).

MATERIALS & METHODS

Retrieval of Amino Acid Sequences

Prolactin amino acid sequences from chimpanzee, sheep, cow and human were obtained from genbank (NCBI). The accession numbers have been shown in Table1.

Table 1: Table showing the accession numbers retrieved from genbank (NCBI).

Species	Prolactin
Cow (*Bos taurus*)	NP_776378
Human (*Homo sapiens*)	NP_000939
Chimp (*Pan troglodytes*)	XP_518264
Sheep (*Ovis Aries*)	NP_001009306

Protein Accession Numbers:

Molecular Modeling

We created protein structures using the Swift Modeler software [6]. We created 10 models for each sequence and selected the one with lowest DOPE score. The selected models were analyzed using VMD software version1.9 (Visual Molecular Dynamics) [7].

Sequence and Structure based alignment

Multiseq (Multiple Sequence Alignment) was used for comparing structures and sequences. Multiseq uses STAMP for structural alignment and Clustal W for sequence based alignment. The aligned models were colored on the basis of their Qres and entropy values for each residue. Phylogenetic analysis was carried out with sequence based (Clustal W) and structure based alignments (STAMP).

Analysis of Amino Acid Composition

Amino acid composition was analyzed using the ExPASy portal in Chimpanzee, Sheep, Human and Cow. Subsequently, we also calculated the molecular weight and theoretical PI for all the four proteins. The results are tabulated (Table 2).

Identification of Functional Motifs

Functional motif was determined by the PROSITE tool in ExPASY. The functional motif is further confirmed by analysis of the 3D structures, using structural similarity determinants such as RMSD and Qres, using VMD software.

RESULT AND DISCUSSION

Structure is primarily responsible for the functional features displayed by a protein. Therefore, structural similarity can be observed in proteins which have a common function. This is because proteins with common functional properties will have common interaction with molecular factors, receptor, etc. The conservation of tertiary structure clearly implies the functional significance of a motif. The functional motif remains conserved during the process of evolution, and experience a selection pressure, which preserves the functional motif.

Prolactin amino acid sequences of chimpanzees, sheep, cow and human were retrieved from genbank (NCBI). These were used to construct protein models using Swift Modeller software. The sequences and structures were aligned using Visual Molecular Dynamics (VMD) software. The protein models were analysed. The structural and sequence based analysis confirmed the presence of two functional motifs in the prolactin gene viz. CHTSSLPTPEDKEQAQQTHHEVLMSLI-LGLLRSW and CLRRDSSKIDTYLKLL-NC.

Bioinformatics tools

From a practical and inexpensive solution for discovering functional motifs in a protein, unlike the labour intensive and costly experimental approaches. Most approaches that are involved in prediction and confirmation of functional motifs involve examining conserved structural motifs, and these are involved in the protein function.

RMSD (Root-mean square deviation) refers to the average distance between atoms of structurally superimposed proteins. The RMSD values for sheep, cow and chimpanzee are very low, as can be observed in the Figure 2. From this we can say that prolactin protein in sheep, cow and chimpanzee possess an overall common 3d structure. The Qres value is another variable to measure the structural similarity between protein structures. It refers to the fraction of native contacts between aligned residues in the two proteins or in two different conformational states of the same protein. The Qres values for cow sheep and chimpanzee were similar. Human prolactin was slightly different as can be observed in Figure2. The Qres and RMSD values clearly indicate the structural similarity as well as the conservation of interatomic distance between the superimposed prolactin protein in cow, sheep and chimpanzee. Human prolactin was the most the most distinct. RMSD (Root-mean square deviation) refers to the average distance between atoms of structurally superimposed proteins.
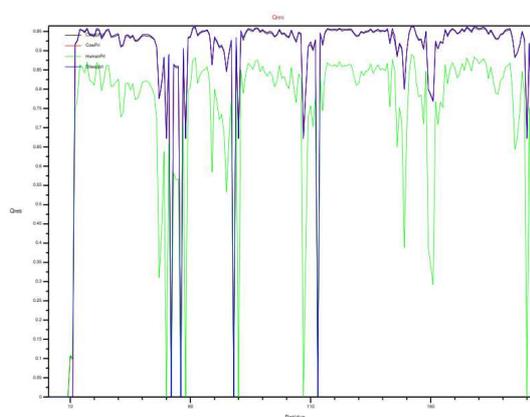

Figure 1: Graphical plot showing Qres values in the aligned of prolactin protein of cow, human, chimpanzee and sheep.
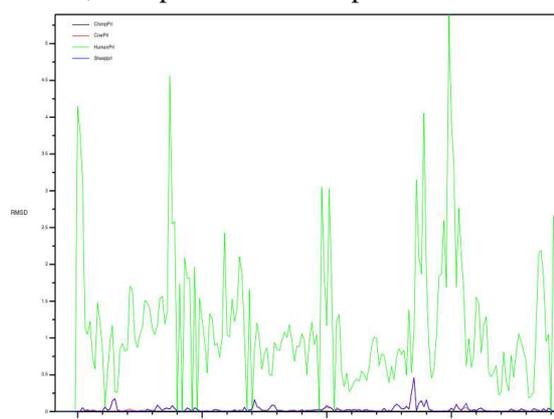

Figure 2: Graphical plot showing RMSD (Root-mean square deviation) values in the Amino Acid Composition (Prolactin) –Expasy Protparam tool
http://web.expasy.org/protparam/

	Human		Cow		Chimp		Sheep	
Ala (A)	12	5.3%	10	4.4%	12	5.3%	10	4.2%
Arg (R)	12	5.3%	12	5.2%	11	4.8%	13	5.4%
Asn (N)	12	5.3%	12	5.2%	12	5.3%	13	5.4%
Asp (D)	8	3.5%	12	5.2%	8	3.5%	12	5.0%
Cys (C)	7	3.1%	7	3.1%	7	3.1%	7	2.9%
Gln (Q)	10	4.4%	9	3.9%	9	4.0%	10	4.2%
Glu (E)	18	7.9%	15	6.6%	18	7.9%	15	6.2%
Gly (G)	9	4.0%	14	6.1%	10	4.4%	14	5.8%
His (H)	9	4.0%	7	3.1%	10	4.4%	9	3.8%
Ile (I)	14	6.2%	11	4.8%	15	6.6%	12	5.0%
Leu (L)	32	14.1%	31	13.5%	32	14.1%	32	13.3%
Lys (K)	12	5.3%	11	4.8%	12	5.3%	11	4.6%
Met (M)	6	2.6%	8	3.5%	5	2.2%	9	3.8%
Phe (F)	5	2.2%	6	2.6%	5	2.2%	6	2.5%
Pro (P)	10	4.4%	11	4.8%	10	4.4%	11	4.6%
Ser (S)	21	9.3%	21	9.2%	21	9.3%	21	8.8%
Thr (T)	8	3.5%	9	3.9%	8	3.5%	10	4.2%
Trp (W)	3	1.3%	2	0.9%	3	1.3%	2	0.8%
Tyr (Y)	7	3.1%	8	3.5%	7	3.1%	7	2.9%
Val (V)	12	5.3%	13	5.7%	12	5.3%	16	6.7%
Pyl (O)	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Sec (U)	0	0.0%	0	0.0%	0	0.0%	0	0.0%
	Number of amino acids: 227 Molecular weight: 25875.8 Theoretical pI: 6.50		Number of amino acids: 229 Molecular weight: 25792.6 Theoretical pI: 5.98		Number of amino acids: 227 Molecular weight: 25767.6 Theoretical pI: 6.36		Number of amino acids: 240 Molecular weight: 27058.1 Theoretical pI: 6.30	

Table 1: Table showing amino acid composition Chimpanzee, Sheep, Human and Cow. Subsequently, we also calculated the molecular weight and theoretical PI for all the four proteins.
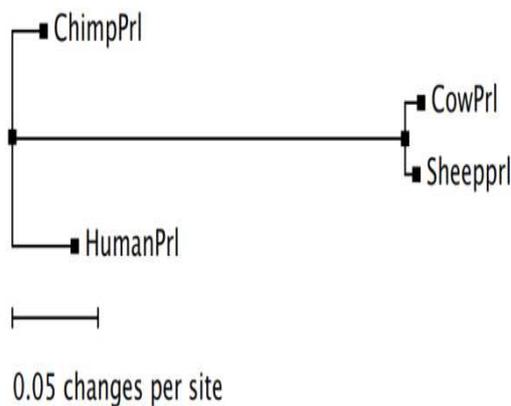

Figure 3: Phylogenetic tree constructed using Clustal W in Multiseq, for sequence based phylogenetic analysis
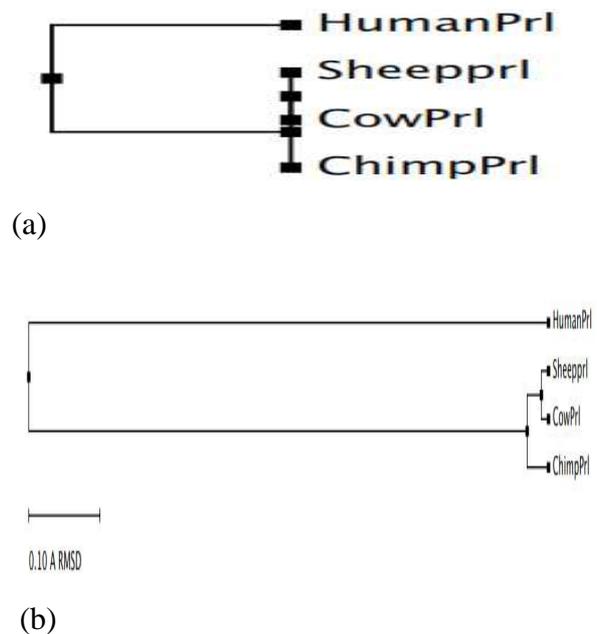

(a)

(b)

Figure 4: Phylogenetic tree constructed using in Multiseq (STAMP alignment), for structure based phylogenetic analysis using (a)Qres values and (b)RMSD values

Phylogenetic analysis based on sequence shows that amino acid sequence of human and chimpanzees prolactin have undergone less change, and are closer, when compared with sheep and cow, which have a higher similarity in the sequences (Fig 3) .

Structure based phylogenetic analysis is in concurrence with RMSD and Qres graph plot results, and consider the human prolactin to be structurally distance than the prolactin from the other three mammals (Fig 4).

ACKNOWLEDGEMENT

REFERENCES

1. Bole-Feysot C, Goffin V, Edery M, Binart N, Kelly PA. Prolactin (PRL) and its receptor: actions, signal transduction pathways and phenotypes observed in PRL receptor knockout mice. *Endocrine Reviews*. 1998: 19 (3):225-68.

2. S. Sharifi, H. Roshanfekr, S.R. Khatami, K.H. Mirzadeh. Prolactin Genotyping of Najdi Cattle Breed Using PCR-RFLP. *Journal of Animal and Veterinary Advances*. 2010; 9 (2):281-283.

3. Le Provost, F., C. Leroux, P. Martin, P. Gaye, J.Dijane. Prolactin gene expression in ovine and caprine mammary gland. *Neuroendocrinology*. 1994; 60: 305-313.

4. Ghasemi N, Zadehrahmani M, Rahimi G, Hafezian SH. Associations between prolactin gene polymorphism and milk production in montebeliard cows. *International Journal of Genetics and Molecular Biology*.2009; 1(3):048-051.

5. Verma S, Agastian P. Expression profile of important milk producing genes in cattle breeds from India. *International Journal of Advanced Biotechnology and Research*. 2013: 1 4 (1): 976-980.

6. Mathur A, Shankaracharya, Vidyarthi AS. SWIFT MODELLER: A JAVA based Graphical User Interface for molecular modeling. *Journal of Molecular Model*. 2010.

7. Humphrey, W., Dalke, A. and Schulten, K. VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics*. 1996: 14: 33-38.