

Computational identification of fertility functions of bovine *Reprimo* gene

Durosaro[1*], S.O., Peters[2], S.O., Adebambo[1], A.O., Onagbesan[4], O.M., Sanda[1], A. J., Olowofeso[1], O. Osho[3], S.O. and Ozoje[1], M.O.
[1]Department of Animal Breeding and Genetics, Federal University of Agriculture, Abeokuta
[2]Berry College, Mount Berry, Georgia Atlanta, U.S.A.
[3]Department of Animal Nutrition, Federal University of Agriculture, Abeokuta
[4]Department of Animal Physiology, Federal University of Agriculture, Abeokuta
*corresponding author: durosaroolutunde@yahoo.com

Abstract

Improvement in fertility is possible through gene assisted selection. Cattle fertility and genes underlying it should be thoroughly studied and exploited to find solution to declining cattle fertility. Reprimo (RPRM) gene is a pleiotropic gene involved in suppression of cancer, regulation of mitotic cell cycle, cell cycle arrest and regulation of survival. Comparison of protein tertiary structures is important in inferring functional characteristics of new proteins. This study used computational approach to identify some fertility functions of bovine RPRM gene using motif prediction and protein structure comparison. Amino acid sequences of bovine RPRM gene and some other cattle fertility genes were retrieved from GenBank. Motifs in the amino acid sequence of bovine RPRM gene were predicted using PROSITE software. The domain structure of bovine RPRM protein was predicted using simple modular architecture research tool (SMART). Protein tertiary structures (3D structures) of bovine RPRM gene and other cattle fertility genes were predicted with Phyre^2 software. To have structural and functional similarity, it has been found that protein structure after superimposition should have Root Mean Square Deviation (RMSD) value less than or equal to 2Å. The protein 3D structures of other cattle fertility genes were superposed against the protein 3D structure of bovine RPRM gene using SuperPose web server and the proteins with RMSD value of 2Å or less were predicted as proteins with similar functions and structures as bovine RPRM gene. The predicted motifs (N-glycosylation site, N-myristoylation site, and cAMP and cGMP-dependent protein kinase phosphorylation site) and protein structure comparison revealed that, bovine RPRM gene and bovine growth hormone gene have the same fertility function with alpha carbon and backbone root mean square deviations of 1.94 Å and 1.81 Å, respectively. It follows therefore that other fertility functions of bovine RPRM gene included sexual maturation, steroidogenesis, gametogenesis, gonadal differentiation and gonadotrophin secretion which are the functions of growth hormone gene.

Keywords: *Reprimo* gene, fertility, growth hormone gene, motifs, root mean square deviation

Introduction

The reproductive performance of livestock is essentially important because it is one of the main determinants of productivity and sustainability of the production system. Successful reproduction requires success in factors connected to fertility. Importance of fertility in cattle is well known, both in functionality and farm economy (Pryce *et al*., 2004).

Animal production demands improvement in functional traits such as pregnancy, stayability, rebreeding rate and calving ease. Fertility is a complicated process, encoded by many genes organized with the frames of functional networks and regulated by many interactions. Therefore, cattle fertility and genes underlying it should be thoroughly studied and exploited to find solution to declining fertility of cattle. Fertility as a complex feature is under the influence of numerous genes, working together to produce functional gametes; promote early embryonic and foetal development and finally the delivery of a healthy calf (Nino-Soto and King, 2004). Despite the low heritabilities of most reproductive traits, reproductive function is under genetic control. Improvements in fertility are possible by selection for quantitative trait loci associated with fertility (Coyral-Castel *et al*., 2010).

Reprimo is a cytoplasmic protein belonging to a family of molecules controlled by p53 that inhibits cell-cycle progression (Hollstein *et al*., 1991). p53, the tumour suppressor gene, is the most commonly mutated gene in human cancer (Levine *et al*., 2004). In healthy cells, upon exposure to genotoxic agents or other noxious particles and stresses, the p53 is activated, resulting in abrogation of the cell cycle (el-Deiry, 1998). This arrest in growth allows for coordination of cellular repair mechanisms and permits the organism to eliminate damaged cells (el-Deiry, 1998). The function of p53 is mediated primarily through activation of target genes (Taylor and Stark, 2001). Previous research has shown that expression of *reprimo* is dependent on p53 (Casson *et al*., 1991) and that over expression of *reprimo* leads to arrest at the G_2 phase of the cell cycle (Ohki *et al*., 2000). *RPRM* gene has also been found by Xu *et al*. (2012) to regulate survival in human. *RPRM* gene is one of the genes enriched in the epididymis and efferent ducts of mouse throughout development (Snyder *et al*., 2010).

Comparison of protein 3D structures is the most fundamental and important task in structural bioinformatics (Zhang and Kim, 2003). Protein structure comparison can be used for various purposes which include analysis of conformational changes on ligand binding, detection of distant evolutionary relationships, inferring functional characteristics of new proteins. Others include assigning folds to new proteins, analysis of structural variation in protein families, assessment of sequence alignment methods, evaluation of structural prediction methods and identification of common structural motifs (Godzik, 1996). Motifs are few amino acids that are critical to a certain function.

Researchers typically solve the structural comparison problem by means of structural alignment or superposition, following the concept of linear sequence alignment (Zhang and Godzik, 2002). Superposition is the process of rotating or orienting an object until it can be directly overlaid on top of a similar object (Maiti *et al*., 2004). Researchers try to find a

maximal set of corresponding pairs (alignment) of amino acid residues that gives a good structural match when superimposed or superposed together. Thus, the terms comparison, alignment, superposition and superimposition are often used interchangeably. Structural alignment generally implies a global alignment, which aligns two structures in their whole, rather than some fragments or portions of them (local alignment). Structural alignment typically means a sequence-order dependent alignment, where the aligned residues must observe the amino acid sequence orders (from N to C-terminus) of two proteins, like in the case of linear sequence alignment. The structural similarity of the two sets of aligned residues is determined by a fitness criterion, typically the Root Mean Square Deviation measures. Root Mean Square Deviation is the measure of the average distance between the atoms (usually the backbone atoms) of superimposed proteins. The general goodness or accuracy of an alignment can be evaluated by a number of quality criteria, each of which is primarily based on the fitness (RMSD), length (number of pairs), and sometimes the number of gaps of the alignment (Kolodyn *et al*., 2005).

The objective of this study was to use motif prediction and protein structure comparison approach to establish the role of *RPRM* gene in cattle fertility.

Materials and Methods

Retrieval of amino acid sequences

Amino acid sequences of *RPRM* gene and some other fertility genes belonging to cattle (*Bos taurus*) were downloaded from GenBank (National Centre for Biotechnology Information database, United States of America). The other cattle fertility genes (Table 1) that were used for protein structure comparison are: Insulin-like growth factor 1 gene (*IGF-1*), Insulin-like growth factor 1 receptor gene (*IGF-1R*), Inhibin beta-A gene (*INHBA*), Signal transducer and activator of transcription 5A gene (*STAT5A*), Growth hormone gene (*GH*), Fibroblast growth factor 2 gene (*FGF2*), Growth hormone receptor gene (*GHR*), Follicle stimulating hormone receptor gene (*FSHR*), Leptin gene (*LEP*), Progesterone receptor gene (*PGR*), Prolactin receptor gene (*PRLP*) and Pregnancy-associated glycoprotein-1 gene (*PAG-1*).

Prediction of protein motifs and domain structure

Motifs in the amino acid sequence of bovine *RPRM* gene were predicted using PROSITE software (De Castro *et al*., 2006). The domain structure of bovine *RPRM* protein was predicted using SMART (Letunic *et al*., 2009).

Prediction and comparison of protein structures

Protein tertiary structures of bovine *RPRM* gene and other cow fertility genes were predicted based on the canonical amino acid sequences obtained from GenBank. The tertiary structures were predicted with Phyre[2]

Table 1: Name, symbol, amino acid sequence length and accession number of bovine *RPRM* and some selected cattle fertility genes for protein structure comparison

Gene name and symbol	Amino acid sequence length	Accession number
Bovine *reprimo* gene (*RPRM*)	109	NP_001074208.1
Pregnancy-associated glycoprotein-1 gene (*PAG-1*)	380	AAB35845.1
Leptin (*LEP*)	167	CAD54745.1
Insulin-like growth factor 1 receptor gene (*IGF-1R*)	640	CAA38724.1
Growth hormone gene (*GH*)	187	CBN80512.1
Growth hormone receptor gene (*GHR*)	300	AAZ15292.1
Fibroblast growth factor 2 gene (*FGF2*)	155	P03969.1
Inhibin beta-A gene (*INHBA*)	425	AAI26557.1
Insulin-like growth factor 1 gene (*IGF-1*)	154	AAI26803.1
Progesterone receptor gene (*PGR*)	383	CAD89880.1
Follicle stimulating hormone receptor gene (*FSHR*)	410	ABV45403.1
Prolactin receptor gene (*PRLP*)	581	AAA51417.1
Signal transducer and activator of transcription 5A gene (*STAT5A*)	794	NP_001012691.1

software (Kelley and Sternberg, 2009). To have structural similarity, it has been found that protein structure after superimposition should have RMSD value less than or equal to 2Å (Shukla *et al*., 2012). The protein tertiary structures of other cattle fertility genes were superposed against the protein tertiary structure of bovine *RPRM* gene using SuperPose web server (Maiti *et al*., 2004) and the protein with RMSD value of 2Å or less was predicted as protein with similar function and structure as bovine *RPRM* gene.

Results

Predicted motifs and domain in bovine RPRM protein

The motifs predicted in bovine *RPRM* protein are shown in Table 2. Three motifs identified in bovine *RPRM* protein were N-glycosylation site at positions 7-10 and 18-21 with consensus sequence (NQTD) and (NSSE) respectively; N-myristoylation site at position 13-18 with consensus sequence (GLFLAN) and cAMP- and cGMP-dependent protein kinase phosphorylation site at position 95-98 with consensus sequence (RRPS).

Only transmembrane region domain was identified in bovine *RPRM* protein. The transmembrane region identified in bovine *RPRM* protein had a consensus sequence (VVQIAVMCVLSLTVVFGIFFLGC) and was located at position 56-78 on the amino acid sequence of bovine *RPRM* protein.

Protein structures of bovine RPRM gene

The modelled tertiary structures of bovine *RPRM* protein are shown in Plates 1-2. Plate 1 illustrates bovine *RPRM* protein in the cartoon model. The alpha helix is depicted as pink coloured spiral sheet, the random coil as blue coloured strand and extended strands as white coloured strands. Plate

2 illustrates the tertiary structure of bovine *RPRM* protein in the space-fill model. The carbon atoms are depicted as light grey balls, nitrogen as blue balls, oxygen as red balls and sulphur as yellow balls.

Comparison of protein 3D structure of bovine RPRM gene with protein 3D structures of other cattle fertility genes

Both alpha carbon RMSD and backbone RMSD were used to ascertain structural similarity between protein 3D structure of bovine *RPRM* gene and protein 3D structures of other selected cattle fertility genes (Table 3). Out of the twelve protein 3D structures of other cattle fertility genes superposed against protein 3D structure of bovine *RPRM* gene, only bovine Growth hormone had significant structural similarity with bovine *RPRM*. Alpha and backbone RMSD values of 1.94 Å and 1.81 Å were observed for the superposition of bovine Growth hormone against bovine *RPRM* respectively. Backbone RMSD of superposition of Insulin-like growth factor-1 receptor against bovine *RPRM* (1.98 Å) was less than 2 Å but the alpha carbon RMSD was more than 2 Å. So, Insulin-like growth factor-1 receptor was not picked as a protein with similar structure with bovine *RPRM* protein. RMSD of all other fertility proteins superposed against bovine *RPRM* protein were greater than 2 Å.

Discussion

The presence of N-glycosylation sites in bovine *RPRM* protein is an indication that the protein is highly glycosylated. Presence of N-glycosylation sites in bovine *RPRM* protein is similar with the findings of Corvalan and Torres (2011) that N-glycosylation sites are also present in human *RPRM* protein. N-glycosylation site was expected in bovine *RPRM* protein because N-glycosylation of proteins is conserved in eukaryotes, and it is one of the most abundant post-translational modification reactions in nearly half of all known proteins in eukaryotes. The presence of asparagine in the N-

Table 3: RMSD values of superposition of protein 3D structures of some cow fertility genes against protein 3D structure of bovine *RPRM* gene

Protein structures superposed	Alpha carbon RMSD (Å)	Backbone RMSD (Å)
PAG-1 against bovine *RPRM*	7.81	7.72
Leptin against bovine *RPRM*	7.57	7.29
IGF-1R against bovine *RPRM*	2.21	1.98
GH against bovine *RPRM*	1.94	1.81
GHR against bovine *RPRM*	6.03	5.90
FGF2 against bovine *RPRM*	9.34	9.28
INHBA against bovine *RPRM*	4.18	3.88
IGF-1 against bovine *RPRM*	6.71	6.53
PGR against bovine *RPRM*	2.35	2.15
FSHR against bovine *RPRM*	2.20	2.14
PRLP against bovine *RPRM*	8.57	8.46
STAT5A against bovine *RPRM*	16.34	16.23

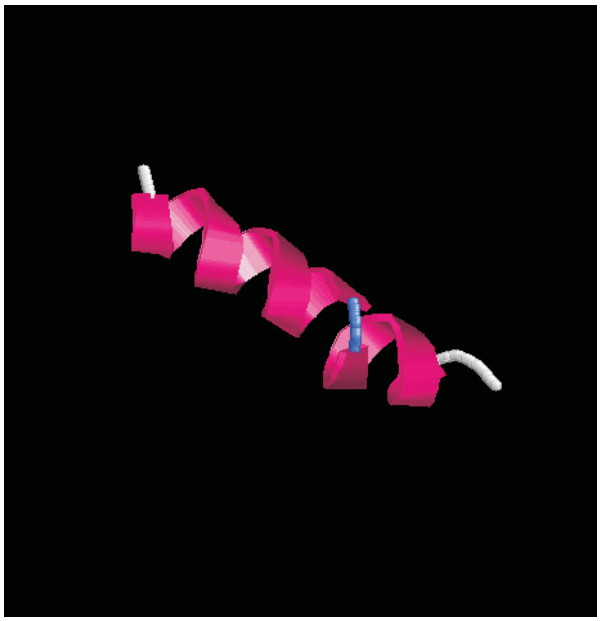

Plate 1: Tertiary structure of bovine *RPRM* protein in cartoon model

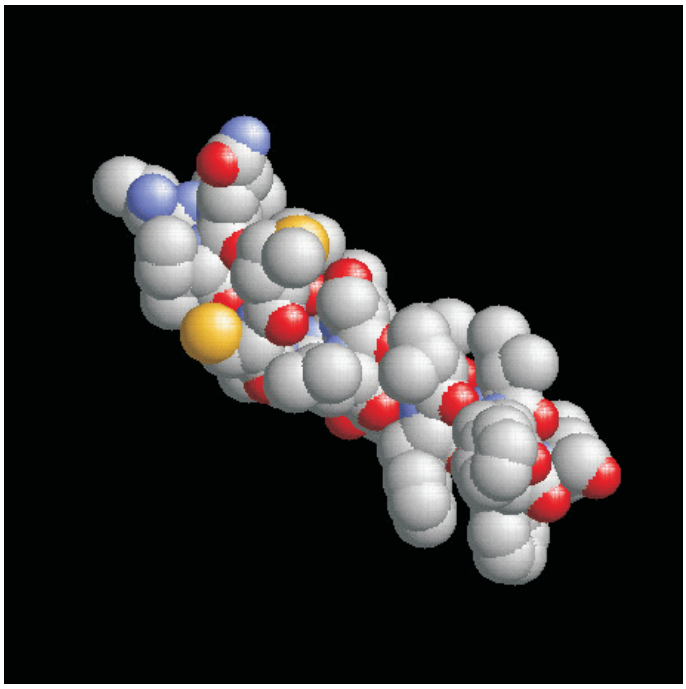

Plate 2: Tertiary structure of bovine *RPRM* protein in space-fill model

glycosylation site of bovine *RPRM* protein is in agreement with previous findings that revealed that protein N-glycosylation requires the presence of asparagine (N) in the consensus sequence (www.worldwidescience.org). N-glycosylation might have also conferred fertility function on bovine *RPRM* gene as Maurer-stroh *et al*. (2002) have reported that N-glycosylation can change the function of modified proteins.

The presence of glycine in myristoylation site of bovine *RPRM* protein is in agreement with the findings of Farazi *et al*. (2001) who reported that myristoylation is more common on glycine residues although it occurs in other amino acids. Presence of myristoylation site in bovine *RPRM* protein will confer fertility function on bovine *RPRM* gene as previous finding by Toraskar *et al*. (2008) has shown that myristoylation is essential for growth, development and rapid cellular proliferation in variety of organisms.

The presence of cAMP and cGMP-dependent protein kinase phosphorylation site in bovine *RPRM* protein will also confer fertility function on the gene as previous work by Jin *et al*. (1999) has shown that cAMP and cGMP are intracellular messengers involved in the transduction of various physiological stimuli and regulation of multiple physiological processes, including visceral motility and reproduction. The cyclic nucleotides cAMP and cGMP play a major role in controlling fertility in various species. Cyclic AMP is also crucially important in regulating the motility of mammalian sperm and the fertilization of the egg (www.caeser.de/index.phpfiid=835&l=2).

The presence of only transmembrane domain in bovine *RPRM* protein implied that the gene function is basically directed by this domain and transmembrane domain is the only folding unit of the protein. This confirms that the single fold observed in the 3D structure model of bovine *RPRM* protein by Phyre[2] is the domain identified in the protein domain prediction. The presence of only one domain in bovine *RPRM* protein might have resulted from small number of amino acid residues in the primary structure of the protein.

The predicted 3D structure of bovine *RPRM* protein in the space-fill model approximately represents the actual shape of the protein in its natural existence while the cartoon model emphasizes its constituent secondary structure elements (Zeyar, 2006). The structural elements observed in bovine *RPRM* protein are responsible for folding, stability and overall functions of the protein. The 3D structure of bovine *RPRM* protein predicted by Phyre[2] was the most likely structure of bovine *RPRM* protein because of the ability of the software to find appropriate protein of known structure as template for the modelling of bovine *RPRM* protein in the protein data bank. Protein 3D structures were used to investigate functional similarities between bovine *RPRM* gene and other

selected cattle fertility genes because the 3D structure is more informative than the linear sequence. Function is more directly correlated with 3D structure than with amino acid sequence (Rasamond and Allsop, 2000). The 3D structure is better conserved than amino acid sequence during evolution (Brenner, 2001). Both alpha carbon and backbone root mean square deviations were used in protein 3D structures comparison because many previous structural comparisons, classifications and clustering applications usually take only the alpha carbon and backbone of proteins into account, and ignore all other atoms for simplicity (Holm and Sander, 1993).

To have structural similarity, it has been found that protein structure after superimposition should have RMSD value less than or equal to 2Å (Shukla *et al*., 2012). Protein structure comparison using RMSD statistics revealed that bovine *RPRM* protein had similar structure with growth hormone. Similarity between the protein 3D structures of bovine *RPRM* gene and growth hormone gene implied that these two genes have similar functions. The comparative protein structure modelling is relevant to structure-based functional annotations of proteins and thus enhances the impact of genome sequencing, structural genomics and functional genomics (Shukla *et al*., 2012). Similarity between the protein 3D structures of both bovine *RPRM* gene and growth hormone gene implied that the gene ontology of bovine *RPRM* gene also includes sexual maturation, steroidogenesis, gametogenesis, gonadal differentiation and gonadotrophin secretion which are the functions of growth hormone gene reported by Zachmann (1992).

Conclusions

Computational prediction of domain structure of bovine *RPRM* protein revealed that only transmembrane domain controls the folding, stability and overall functions of the protein. Protein structure comparison using RMSD statistics revealed that bovine *RPRM* protein had similar structure with growth hormone which implied that bovine *RPRM* gene and growth hormone gene have the same fertility functions which include sexual maturation, steroidogenesis, gametogenesis, gonadal differentiation and gonadotrophin secretion. Therefore bovine *RPRM* gene can be used in place of bovine growth hormone gene in marker assisted selection when screening cattle for fertility

References

Brenner, S.E. 2001. A tour of structural genomics. *Nature Reviews Genetics*. 2(10): 801-809.

Casson, A.G., Mulhopadhyay, T., Cleary, K.R., Ro, J.Y., Levin, B. and Roth, J.A. 1991. p53 mutations in Barrett's epithelium and esophageal cancer. *Cancer Research.* 51: 4495-4499.

Corvalan, A.H. and Torres, V.A. 2011. *RPRM* (*Reprimo,* TP53 dependent G_2 arrest mediator candidate). *Atlas of Genetics, Cytogenetics, Oncology and Haematology*. http://AtlasGeneticsOncology.org/Genes/*RPRM*ID42082ch2q23.html.

Coyral-Castel, S., Rame, C., Monniaux, D., Freret, S., Fabre-Nys, C., Fritz, S., Monget, P., Dupont, F. and Dupont, J. 2010. Ovarian parameters and fertility of dairy cows selected for one QTL located on BTA3. *Theriogenology*. doi: 10.1016/j.theriogenology.11.037.

De Castro, E., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A. and Hulo, N. 2006. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research.* **34:** W362–W365.

el-Deiry, W.S. 1998. Regulation of p53 downstream genes. *Seminar on Cancer Biology*. 8: 345-357.

Farazi, T.A., Waksman, G. and Gordon, J.I. 2001. The biology and enzymology of protein N-myristoylation. *The Journal of Biological Chemistry*. 276(42): 39501-39504.

Godzik, A. 1996. The structural alignment between two proteins: is there a unique answerfi *Protein Science*. 5(7): 1325-1338.

Hollstein, M., Sidransky, D., Vogelstein, B. and Harris, C.C. 1991. p53 mutations in human cancers. *Science*. 253:49-53.

Holm, L. and Sander, C. 1993. Protein structure comparison by alignment of distance matrices. *Journal of Molecular Biology*. 233: 123-138.

Jin, S.K., Richard, F.J., Kuo, W.P., D'Ercole A.J. and Conti, M. 1999, Impaired growth and fertility of cAMP-specific phosphodiesterase PDE4D-deficient mice. *Proceedings of the National Academy of Sci*ences. 96: 11998 – 12003.

Kelley, L.A. and Sternberg, M.J.E. 2009. Protein Structure and Prediction on the web: a case study using the Phyre server. *Nature Protocols*. 4: 363-371.

Kolodny, R., Koehl, P. and Levitt, M. 2005. Comprehensive evaluation of protein structure alignment methods: scoring by geometric measures. *Journal of Molecular Biology*. 346: 1173-1188.

Letunic, I., Doerks, T. and Bork, P. 2009. SMART 6: recent updates and new developments. *Nucleic Acids Research.* **37:** D229–D232.

Levine, A.J., Finlay, C.A. and Hinds, P.W. 2004. p53 is a tumour suppressor gene. *Cell*. 116: S67-S69.

Maiti, R., Van Domselaar, G.H., Zhang, H. and Wishart, D.S. 2004. Superpose: a simple server for sophisticated structural superposition. *Nucleic Acids Research*. 32: W590-W594.

Maurer-Stroh, S., Eisenhaber, B. and Eisenhaber, F. 2002. N-terminal N-myristoylation of proteins: refinement of the sequence motif and its taxon-specific differences. *Journal of Molecular Biology*. 317(4): 523-540.

Nino-Soto, M.I. and King W.A. 2004. Genetic factors that affect normal reproduction and fertility in domestic cattle. *Proceeding of the 23rd World Buiatrics Congress, Quebec, Canada*. Pp 567-70.

Ohki, R., Nemoto, J. and Murasawa, H. 2000. *Reprimo*, a new candidate mediator of the p53-mediated cell cycle arrest at the G_2 phase. *Journal of Biological Chemistry.* 275: 22627-22630.

Pryce, J.E., Royal, M.D., Garnsworthy, P.C. and Mao, I.L. 2004. Fertility in the high-producing cows. *Livestock Production Sci*ence. 86: 125-135.

Rasamond, J. and Allsop, A. 2000. Harnessing the power of the genome in the search for new antibiotics. *Science*. 287(5460): 1973-1976.

Shukla, S., Dixit, S. and Pardasani, K.R. 2012. *In silico* identification of anti-fertility proteins based on sequence and structural similarity. *Asian Journal of plant Science and Research*. 2(3): 290-298.

Synder, E.M., Small, C.L., Bomgardner, D., Xu, B., Evanoff, R., Griswold, M.D. and Hinton, B.T. 2010. Gene expression in the efferent ducts, epididymis and vas deferens during embryonic development of the mouse. *Developmental Dynamics.* 239: 2479-2491.

Taylor, W.R. and Stark, G.R. 2001. Regulation of the G_2-M transition by p53. *Oncogene*. 20: 1803-1815.

Toraskar, M.P., Prasad, K.K. and Kadam, V.J. 2008. N-myristoyltransferase: A novel target. *Mini Reviews in Medicinal Chemistry*. 8(2): 142-149.

www.caeser.de/index.phpfiid=835&l=2. Assessed on 27/05/2013.

www.worldwidescience.org. Assessed on 12/03/2013.

Xu, M., Knox, A.J., Michaelis, K.A., Kiseljak-Vassiliades, K., Kleinschmidt-DeMasters, B.K., Lillehei, K.O. and Wierman, M.E. 2012. *Reprimo* (*RPRM*) is a Novel Tumour Suppressor in Pituitary Tumours and Regulates Survival, proliferation and Tumorigenicity. *Endocrinology en*. 2011-2021.

Zachmann, M. 1992. Interrelations between growth hormone and sex hormones-physiology and therapeutic consequences. *Hormone Research.* 38: 1-8.

Zeyar, A. 2006. *Computational analysis of 3D protein structures*. A thesis submitted for the Degree of Doctor of philosophy, School of Computing, National university of Singapore. 190pp.

Zhang, B. and Godzik, A. 2002. The meaning and limitations of protein structure alignments. *Proceedings of 1st International Symposium on 3D Data Processing Visualization and Transmission*. Pp 726-729.

Zhang, C. and Kim, S. H. 2003. Overview of structural genomics: from structure to function. *Current Opinion in Chemical Biology*. 7: 28-32.

Received: 8th December, 2014
Accepted:22nd December, 2015