Acta Scientific Veterinary Sciences (ISSN: 2582-3183)

Volume 2 Issue 3 March 2020

Research Article

An *In-Silico* Structural Characterization of the Buffalo Steroidogenic Proteins

Emmagouni Sharath Kumar Goud, Mamta Pandey, Prashant Singh, Chhama Singh, Vedamurthy Gowdar Veerappa, Dheer Singh and Suneel Kumar Onteru*

Molecular Endocrinology, Functional Genomics, and Systems Biology Lab, Animal Biochemistry Division, ICAR-National Dairy Research Institute, Karnal, India

***Corresponding Author:** Suneel Kumar Onteru, Senior Scientist, Molecular Endocrinology, Functional Genomics, and Systems Biology Lab, Animal Biochemistry Division, ICAR-National Dairy Research Institute, Karnal, India.

DOI: 10.31080/ASVS.2020.02.0051

Received: February 06, 2020

Published: February 21, 2020

Abstract

Post-partum reproductive disorders are of major concern in buffaloes. Mainly these are regulated by steroidogenic proteins, such as CYP17, CYP19, and 3β-HSD. These enzymes are involved in the synthesis of several steroid hormones, an imbalance in the levels of which can lead to the causation of several reproductive illnesses ultimately affecting milk production. In the present *in silico* study, we analyzed the structural details of the three steroid enzymes in terms of their physiochemical properties, N- and O-glycosylation sites, phosphorylation sites, secondary structure, surface probability, hydrophilicity and antigenic index. Additionally, the 3D models for the three proteins were enacted in the SWISS-MODEL online tool and the models were assessed by RAMPAGE. It was found that the CYP17 protein sequence of buffalo is small having 247 aminoacids when compared with other species. All the three buffalo steroid proteins were having more than 95% similarity with cattle, sheep and goat, except human. The amino acids responsible for the heme-binding site in CYP17, a catalytic site in CYP19, and 3β-HSD proteins in buffalo were found to be conserved when compared with human protein sequences. The 3D models predicted for the three buffalo steroidogenic proteins were found to be of good quality through Ramachandran plot. Further, the individual phylogenetic trees for each of the protein showed that human proteins are phylogenetically outgroup to buffalo proteins. Overall, the *in-silico* analysis of the three buffalo steroidogenic proteins could prove to arise new insights for resolving reproductive disorders in the buffaloes.

Keywords: Steroidogenic; *In-Silico;* Buffalo; Steroid

Introduction

Buffaloes are one of the premier dairy animals in developing countries, especially in South East Asia. For example, India is one of the top countries in milk production with an annual production of 155.5 million tons in 2015-16 [1]. Buffaloes contribute about 55% [2] of the total milk production in India. Although this contribution is prominent to the dairy sector, the milk productivity of buffaloes is not up to the mark. Reproductive problems are one of the reasons for less productivity in buffaloes.

Postpartum anestrus condition, polycystic ovaries, infertility etc., are the major reproduction problems in buffaloes. There can be many reasons for these reproductive problems, one of which is a hormonal imbalance. For instance, cystic ovaries consisting of persistent corpus leuteum produces high levels of progesterone which leads to anestrus condition. Polycystic ovary condition also leads to hyperandrogenism [3]. Considering the molecular insights, these problems may arise due to an improper expression of genes and proteins involved in hormone synthesis pathways. Three such proteins of major importance in steroid hormone, particularly estrogen biosynthesis, are cytochrome P450 17 alpha-hydroxylase/17, 20-lyase (CYP17), 3β-HSD (3β-Hydroxysteroid dehydrogenase/Δ^{5-4} isomerase) and CYP19 (Cytochrome P-450 aromatase).

The CYP17and CYP19 are the members of cytochrome P450 superfamily of enzymes involved in steroid biosynthesis. These proteins are primarily localized in the endoplasmic reticulum (ER) membrane [4]. The CYP17 shows 17α-hydroxylase and 17, 20-lyase activities. The 17α-hydroxylase catalyzes the hydroxylation of pregnenolone and progesterone into 17α-hydroxy pregnenolone and17α-hydroxy progesterone, respectively. The 17, 20-lyase activ-

ity is involved in the conversion of 17α-hydroxy pregnenolone and 17α-hydroxy progesterone to dehydroepiandrosterone and androstenedione, respectively. It was found that mutations in the CYP17 gene leads to a deficiency of 17α-hydroxylase/17, 20-lyase causing pseudo hermaphroditism, adrenal hyperplasia, amenorrhea and the lack of secondary sexual characteristics in women [5]. Amenorrhea or lack of female sexual cycles is a kind of similar condition of true anestrus in buffaloes. Hence, characterization of the CYP17 protein is important even in buffaloes.

Aromatase is the key enzyme for estrogen biosynthesis encoded by the CYP19 gene [6]. Aromatase catalyzes the conversion of androstenedione to estrone and testosterone to estradiol by aromatization at ring A. The catalytic complex of CYP19 includes the aromatase enzyme with a catalytic iron-binding porphyrin ring as a prosthetic group in the active site, and the NADPH–cytochrome P450 reductase enzyme having required amount of NADPH for the reaction [7]. The deficiency of aromatase enzyme results in the low levels of estrogens in the ovaries which could be a reason for silent heat in buffaloes. Silent heat may be a predominant cause of anestrus cases in the field conditions [8]. The lower activity of aromatase results in hyperandrogenism which may lead to anovulation and infertility [9]. However, aromatase protein and its structure is not yet well elucidated in buffaloes.

The 3β-HSD (3β-Hydroxysteroid dehydrogenase/Δ^{5-4} isomerase) is the enzyme involved in the biosynthesis of progesterone from pregnenolone, 17α-hydroxy progesterone from 17α-hydroxy pregnenolone, androstenedione from DHEA in steroid hormone biosynthesis pathways. Its deficiency resulted in lower levels of estrogens, which necessitates the sex hormone therapy at puberty in humans [10]. As late maturity is one of the common problems in buffaloes, understanding the buffalo 3β-HSD at the molecular level is required.

The goal of this manuscript is to analyze the structural details of the three important proteins (CYP17, CYP19, and 3β-HSD) of the steroidogenesis pathway in buffaloes. This preliminary step through computational biology may help to modulate these protein activities in the future for handling the reproductive problems of buffaloes.

Materials and Methods

Sequence retrieval and Multiple sequence alignment (MSA)

The amino acid sequences of the buffalo CYP17, CYP19,and3β-HSD proteins were retrieved from the NCBI protein database. The retrieved buffalo protein sequences were used in the BLASTp online tool to get the protein sequences for other species like cow, goat, sheep, and human. The details of the protein sequences along with their species name, gene name and accession numbers are presented in table 1. The sequences were retrieved in a fasta format and aligned through Clustal Ω online tool with default parameters [11].

Species Name	Protein IDs		
	CYP17	CYP19	3β-HSD
Bubalus bubalis	XP_006065001.1	ABD64825.2	XP_006049419.1
Bos taurus	NP_776729.1	AAA62244.1	NP_776768.1
Bos indicus	XP_019844145.1	XP_019824386.1	XP_019812272.1
Ovis aries	XP_011958253.1	NP_001116472.1	NP_001129404.1
Capra hircus	ABQ12616.1	NP_001272676.1	NP_001272645.1
Homo sapiens	AAA59984.1	AAH35959.1	NP_000853.1

Table 1: Accession numbers of the steroidogenic proteins.

Primary protein sequence analysis

The physiochemical properties such as molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) were computed for the buffalo CYP17, CYP19, and 3β-HSD protein sequences by online tool Protparam [12,13]. The N-Glycosylation, O-Glycosylation, and Phosphorylation sites of the these proteins were predicted using the online tools NetOGlyc 3.1 server [14,15], NetNGlyc 1.0 server [16,17] and NetPhos 3.1 server [18,19], respectively. The signal peptide and transmembrane helices were predicted from the protein sequences using the SOSUI signal online software [20,21].

Secondary and tertiary structure prediction

The Protean tool (DNASTAR) was used for secondary structure prediction [22], hydrophilicity plot [23], antigenic index [24] and surface probability plot (Emini). The tertiary structure was predicted using the online software SWISS-MODEL [25,26]. The protein structure obtained from SWISS-MODEL was subjected to Ramachandran's plot analysis using the RAMPAGE online tool [27].

Construction of phylogenetic tree

The evolutionary tree was constructed by the MEGA7 software [28] using maximum likelihood method on the basis of Jones-Taylor-Thornton (JTT) matrix-based model [29]. For initial inference of the tree, the pairwise distances estimated by the JTT model between proteins were used for the heuristic search by applying the

Nearest-Neighbour Interchange (NNI) and Bio NJ algorithms. The reliability of the branching of the tree was checked by 1000 bootstrap samplings [30]. The branches corresponding to the partitions were collapsed when they were reproduced in less than 50% of the bootstrap replicates. The gaps and missing data at all positions were eliminated.

Ethical approval

The present study is an *in silico* study utilizing the available online gene sequences and software. It does not involve any animal studies. Hence, the study does not require any animal ethics committee approval.

Results

Multiple sequence alignment

The MSA for the retrieved 247 amino acids (a.a) of buffalo CYP17 steroid protein showed 97.98% similarity with *Bos taurus,* 96.76% with *B. indicus, Ovis aries* and *Capra hircus*, and 71.95% with that of *Homo sapiens* (Figure 1a). The buffalo CYP19 protein sequence showed approximately 98% similarity with all the above mentioned ruminant species and 83% with that of human (Figure 1b). Similarly, the buffalo 3β-HSD protein has 99% identity with *B. taurus* and *B. indicus*, 97% identity with *O. aries* and *C. hircus*, and 79% with *H. sapiens* sequence (Figure 1c).

Figure 1

Protein sequence analysis

The number of negatively (Asp+Glu) and positively (Arg+Lys) charged amino acid residues were 20 and 11 for the CYP17, 61 and

S. No	Properties	CYP17	CYP19	3β-HSD
1.	No. of amino acids	247	503	373
2.	Molecular weight	27999.9	58071.27	42272.55
3.	Theoretical Pi	9.86	6.87	7.17
4.	Amino acid composition (%)			
	Ala (A)	5.3	4.4	5.9
	Arg (R)	5.3	5.6	3.8
	Asn(N)	3.6	4.0	4.6
	Asp (D)	2.0	5.2	3.5
	Cys (C)	1.2	2.2	2.4
	Gln (Q)	5.3	2.4	4.0
	Glu (E)	6.1	7.0	7.0
	Gly (G)	6.5	5.0	6.4
	His (H)	2.0	2.4	3.2
	Ile (I)	4.9	8.0	5.9
	Leu (L)	14.6	10.9	11.3
	Lys (K)	8.9	6.4	6.7
	Met (M)	2.4	5.0	1.3
	Phe (F)	6.1	5.8	3.5
	Pro (P)	6.5	4.2	4.3
	Ser (S)	8.1	5.8	7.2
	Thr (T)	3.2	3.8	5.1
	Trp (W)	1.2	1.2	2.1
	Tyr (Y)	2.0	3.2	4.8
	Val (V)	4.9	8.0	7.0
5.	Total number of negatively charged residues (Asp + Glu)	20	61	39
6.	Total number of positively charged residues (Arg + Lys)	35	60	39
7.	Atomic composition:			
	Carbon (C)	1279	2634	1915
	Hydrogen (H)	2041	4157	2972
	Nitrogen (N)	343	681	504
	Oxygen (O)	343	722	548
	Sulfur (S)	9	36	14
8.	Formula	$C_{1279}H_{2041}N_{343}O_{343}S_9$	$C_{2634}H_{4157}N_{681}O_{722}S_{36}$	$C_{1915}H_{2972}N_{504}O_{548}S_{14}$
9.	Total number of atoms	4015	8230	5953
10.	Extinction coefficients (M^{-1} cm^{-1}, at 280 nm)	24075	57465	71320
11.	Estimated half-life			
(a)	Mammalian reticulocytes, *in vitro*	30 hours	30 hours	30 hours

(b)	Yeast, *in vivo*	>20 hours	>20 hours	>20 hours
(c)	*Escherichia coli, in vivo*	>10 hours	>10 hours	>10 hours
12.	Instability index (II)	41.39	32.42	43.57
13.	Aliphatic index	95.14	101.09	93.03
14.	Grand average of hydropathicity (GRAVY)	-0.181	0.063	-0.194

Table 2: *In-silico* analysis for the physicochemical properties of buffalo steroid biosynthesis proteins.

60 for the CYP19 and 39 and 39 for the 3β-HSD. The total number of negatively charged residues (Asp+Glu) and positively charged residues (Arg+Lys) were known to indicate the nature of proteins. Accordingly, the CYP17 is basic in nature with an overall positive charge, and the CYP19 and 3β-HSD proteins are neutral in nature. The other properties like molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, grand average of hydropathicity (GRAVY) etc., were also computed as shown in table 2. The instability index (32.42), aliphatic index (101.09) and GRAVY score (0.063) indicated that buffalo CYP19 had higher stability in a test tube (*in vitro*), thermostability and hydrophobicity than the CYP17 and 3β-HSD.

The analysis of O-glycosylation, N-glycosylation, and Phosphorylation sites revealed the absence of significant O-glycosylation sites in all the three proteins. However, the CYP17, CYP19, and 3β-HSD have only one N-glycosylation site at Asn217, Asn12, and Asn268, respectively. Similarly, the CYP17, CYP19, and 3β-HSD have 14, 33 and 34 phosphorylation sites, respectively.

By using the SOSUI signal software, we predicted that CYP17 protein consisted of one signal peptide but lacks transmembrane helices, the CYP19 consisted of one transmembrane helix but lacks a signal peptide, and the3β-HSD consisted of both signal peptide and transmembrane helix. The signal peptide of CYP17 is of 17 a.a. in length present at the N-terminal end starting from the 1st-17th position. The transmembrane helix of the CYP19 protein was found to be of 23 a.a length present at the N-terminal end starting from 17th–39th position. The signal peptide and transmembrane helix of 3β-HSD were of 25a.a's from 1st - 25th position and 23 a.a's length from the 289th - 311th position, respectively.

Secondary and tertiary structure prediction

Upon comparative secondary structure analysis, it was observed thatCYP19 protein has more alpha helices and beta sheets than CYP17 and3β-HSD protein sequences, whereas the turns are more in number in CYP17 and coils are more in CYP17 and 3β-HSD than CYP19. The antigenicity was found to be higher for CYP19 followed by 3β-HSD and CYP17 on the basis of a number of antigenic peptides (Figure 2).

Figure 2

The tertiary structures of the three buffalo steroid proteins were predicted through the SWISS-MODEL (Figure 3). The tertiary structures thus obtained were submitted to RAMPAGE in PDB format for the validation of predicted protein structures. From the Ramachandran plot, it was analyzed that CYP17 protein has 96.2% of residues in the favored region, 2.4% in the allowed region and 1.4% in the outlier region. The CYP19 protein has 94% in the favored region, 5.8% in the allowed region and 0.2% in the outlier region. The 3β-HSD protein had 89.3% in the favored region, 8.6% in the allowed region and 2.1% in the outlier region (Figure 4).

Figure 3

Construction of phylogenetic tree

The phylogenetic tree constructed using MEGA7 software depicted that the CYP17 protein of *Bubalus* was phylogenetically more similar to both *Bos taurus* and *Bos indicus* proteins than

Figure 4

other species sequences. CYP19 protein of buffalo was more phylogenetically similar to small ruminant species i.e. Ovis and Capra than large ruminants and human sequences. However, the buffalo 3β-HSD and CYP17 proteins were found to be more similar to *Bos taurus* and *Bos indicus* than small ruminants and humans (Figure 5).

Figure 5

Conclusion

The present study provides the *in-silico* analysis and structural models of the three buffalo proteins which are involved in steroidogenesis i.e. CYP17, CYP19, and 3β-HSD. These proteins play a crucial role in the metabolism of cholesterol towards the synthesis of various steroid hormones, especially of those involved in reproduction [31]. Considering the importance of these three enzymes or proteins in steroidogenesis, and their structural details are unknown for buffaloes, it becomes necessary to know their properties and structure for buffaloes. A single defect in any of these three proteins could affect the reproductive cycle of buffaloes, leading to a decrease in their reproductive efficiency, thus productivity.

The CYP17 buffalo protein sequence, as predicted by NCBI, is smaller (247 a.a) than the protein sequences of other species, which are of nearly 500 a.a. The heme-binding site of CYP17 was found at 171-191 position in buffalo protein sequence. This sequence of heme binding site is present at a 433 - 453 position in the human CYP17 protein sequence [32]. Sequence alignment of buffalo and human CYP17 proteins inferred that the conserved positively charged arginine residue having selective lyase activity is present at the 187th position in buffalo sequence instead of 449th position [33] of the human and other species sequence.

The amino acid number for the CYP19 was observed to be the same for all the species (503 a.a) used for comparison in this study. The amino acid residues present at positions i.e. F134, F221, W224, I305, A306, D309, T310, V370, L372, V373, M374, L477, and S478 were contributing to the catalytic cleft of human aromatase [34] where androstenedione precisely fits in the cleft. Upon comparison of the residues, we can state that the catalytic cleft residues of the buffalo and human aromatase are exactly same and conserved for a CYP19 protein sequence. Hence, it may be possible to use inhibitors (e.g., letrazol) or activators (e.g.., zeranol) of human aromatase to modulate the activity of buffalo aromatase. Future *in vitro* and *in vivo* studies are required to understand the effect of human aromatase inhibitors in buffaloes to mitigate the reproductive problems.

In the present study, buffalo 3β-HSD protein sequence was found to be 79% similar to human 3β-HSD protein sequence. This finding was in accordance with the reports of Zhao., *et al.* [35], who also observed a similar identity for bovine and human 3β-HSD protein sequence. 3β-HSD protein is of two types i.e. type1 3β-HSD and type2 3β-HSD. The difference is for the presence of the type of amino acid at the 187th amino acid residue which is methionine for type1 3β-HSD and threonine for type2 3β-HSD [36]. By analyzing the amino acid composition of the buffalo 3β-HSD protein in the present work, it can be inferred that buffalo 3β-HSD protein sequence is of type 13β –HSD as methionine is present at a 187th position in the sequence. In buffalo 3β-HSD, we found Glu142 and Tyr253 residues, which are believed to have a crucial role in its catalytic activity [37]. Also, Glu142 and Tyr253 are highly conserved in many species as well as in both 3β-HSD isoforms [37]. The key residues, Asn100 and Glu126, were also conserved in buffalo 3β-HSD proteins. The Asn100 and Glu126in human 3β-HSD1 are involved in the dehydrogenase and isomerase activity, respectively [38].

The protein CYP17 and CYP19 are localized in the endoplasmic reticulum where the pH is 7.2. The theoretical pI for the CYP17 is 9.86 and it has a number of positively charged amino acids, indicating its activity in the basic environment. Similarly, the theoretical pI of 3β-HSD was found to be 7.17, signifying the stability of this protein in a basic environment. 3β-HSD is known to be localized in microsomes and mitochondria where pH is 8.0. The CYP19 having a theoretical pI of 6.87 showed its near stable subcellular localization in the ER. An instability index greater than 40 is an indication

of instability of the protein in a test tube environment [39]. Considering the same, the buffalo CYP17 and 3β-HSD are unstable while CYP19 is stable in a test tube system.

N-linked glycosylations are universal modifications present in all eukaryotes [40] that can modify thermodynamic, kinetic and structural properties of the proteins depending on their sequence [41]. The prediction of N-glycosylation sites in buffalo CYP19 protein showed glycosylation at Asn12. The N-glycosylation at this specific site helps in subcellular localization of CYP19 protein into the lumen of ER (Kaur and Bose 2014). Similar to our study in the buffalo, the Asn12 site is also present in human aromatase [42]. In our study, N-glycosylation site (Asn-X-Thr) was located at amino acids 268^{th} - 270^{th} positions of 3β-HSD protein in Buffalo. This finding was similar to the reports of 3β-HSD protein in bovine [35], indicating its conserved function between buffaloes and cows.

In the present study, we noticed the absence of signal peptide in buffalo aromatase protein similar to human aromatase. Although the signal peptide is absent, the human aromatase protein was found to be translocated by signal anchor-mediated translocation [43]. A similar translocation mechanism may be expected for buffalo aromatase protein. In our study, the antigenic index, hydrophilicity, and surface probability were observed to have a correlated pattern for all the three steroidogenic proteins. The regions showing hydrophilic domains in the hydrophilicity plot showed higher antigenic index for all the three proteins. It was concluded that the regions with higher hydrophilicity and surface probability show higher antigenic index.

The results as obtained from Ramachandran plot had indicated that the maximum amino acid residues for all the three protein sequences were found to be present in the favored and allowed region indicating an overall stable protein configuration of the predicted structures. From the phylogenetic trees, we found that all the three proteins are well conserved in buffalo throughout the course of evolution.

Conclusion

In this study, the structural details of buffalo steroidogenic enzymes were characterized by *in silico* approaches. Particularly, physicochemical properties, glycosylation sites, phosphorylation sites, signal peptide, transmembrane helices, secondary and tertiary structure, antigenic index and a phylogenetic tree were identified for the three steroidogenic proteins. In the future, these structural details will be useful for designing drugs to target these proteins for therapeutics of buffalo reproductive disorders.

Conflict of Interest

The authors declare that they do not have any conflict of interests.

Bibliography

1. http://www.nddb.org/information/stats/milkprodindiaz
2. http://www.milkproduction.com/Library/Editorial-articles/Milk-quality-in-India/
3. Puurunen J., *et al*. "Adrenal androgen production capacity remains high up to menopause in women with polycystic ovary syndrome". *The Journal of Clinical Endocrinology and Metabolism* 94.6 (2009): 1973-1978.
4. Neve EP and Ingelman-Sundberg M. "Cytochrome P450 proteins: retention and distribution from the endoplasmic reticulum". *Current Opinion in Drug Discovery and Development* 13.1 (2010): 78-85.
5. Fernández-Cancio M., *et al*. "Discordant genotypic sex and phenotype variations in two Spanish siblings with 17α-hydroxylase/17, 20-lyase deficiency carrying the most prevalent mutated CYP17A1 alleles of Brazilian patients". *Sexual Development* 11.2 (2017): 70-77.
6. Kumar OS., *et al*. "CYP19 (cytochrome P450 aromatase) gene polymorphism in murrah buffalo heifers of different fertility performance". *Research in Veterinary Science* 86.3 (2009): 427-437.
7. Poulos TL., *et al*. "The 2.6-A crystal structure of Pseudomonas putida cytochrome P-450". *Journal of Biological Chemistry* 260.30 (1985): 16122-16130.
8. Hafez ESE and Hafez B. "Reproduction in farm animals". Lippincott Williams and Wilkins publications, (7ed) (2000): 261-263.
9. Chen J., *et al*. "The correlation of aromatase activity and obesity in women with or without polycystic ovary syndrome". *Journal of Ovarian Research* 8.1 (2015): 11.
10. Kolon TF., *et al*. "Disorders of sexual development". *Penn Clinical Manual of Urology*, *Elsevier Health Sciences* (2007): 827-852.
11. Sievers F., *et al*. "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega". *Molecular Systems Biology* 7.1 (2011): 539.

12. http://web.expasy.org/protparam/

13. Gasteiger, E., *et al*. "Protein identification and analysis tools on the ExPASy server". *Humana Press* (2005): 571-607.

14. http://www.cbs.dtu.dk/services/NetOGlyc-3.1/

15. Julenius K., *et al*. "Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites". *Glycobiology* 15.2 (2005): 153-164.

16. http://www.cbs.dtu.dk/services/NetNGlyc/

17. Gupta R., *et al*. "Prediction of N-glycosylation sites in human proteins". 46 (2004): 203-206.

18. http://www.cbs.dtu.dk/services/NetPhos/

19. Blom, N., *et al*. "Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence". *Proteomics* 4.6 (2004): 1633-1649.

20. http://harrier.nagahama-i-bio.ac.jp/sosui/sosuisignal/sosuisignal_submit.html

21. Gomi M., *et al*. "High performance system for signal peptide prediction: SOSUI signal". *Journal of chem-bio informatics* 4.4 (2004): 142-147.

22. Garnier J., *et al*. "GOR method for predicting protein secondary structure from amino acid sequence". *Methods in Enzymology* 266.32 (1996): 540-553.

23. Kyte J and Doolittle RF. "A simple method for displaying the hydropathic character of a protein". *Journal of Molecular Biology* 157.1 (1982): 105-132.

24. Jameson BA and Wolf H. "The antigenic index: a novel algorithm for predicting antigenic determinants". *Bioinformatics* 4.1 (1988): 181-186.

25. https://swissmodel.expasy.org/

26. Guex N., *et al*. "Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective". *Electrophoresis* 30.S1 (2009).

27. http://mordred.bioc.cam.ac.uk/~rapper/rampage.php

28. Kumar S., *et al*. "MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets". *Molecular Biology and Evolution* 33.7 (2016): 1870-1874.

29. Jones DT., *et al*. "The rapid generation of mutation data matrices from protein sequences". *Bioinformatics* 8.3 (1992): 275-282.

30. Felsenstein J. "Confidence limits on phylogenies: an approach using the bootstrap". *Evolution* 39.4 (1985): 783-791.

31. Hu J., *et al*. "Cellular cholesterol delivery, intracellular processing and utilization for biosynthesis of steroid hormones". *Nutrition and metabolism* 7.1 (2010): 47.

32. Picado-Leonard., *et al*. "Cloning and sequence of the human gene for P450cl7 (steroid) 17α-hydroxylase/17, 20 lyase: similarity with the gene for P450c21". Dna 6.5 (1987): 439-448.

33. Akhatar MK., *et al*. "Cytochrome b5 modulation of 17α hydroxylase and 17-20 lyase (CYP17) activities in steroidogenesis". *Journal of Endocrinology* 187 (2005): 267-274.

34. Ghosh D., *et al*. "X-ray structure of human aromatase reveals an androgen-specific active site". *The Journal of Steroid Biochemistry and Molecular Biology* 118.4-5 (2010): 197-202.

35. Zhao HF., *et al*. "Molecular cloning, cDNA structure and predicted amino acid sequence of bovine 3β-hydroxy-5-ene steroid dehydrogenase/Δ5-Δ4 isomerase". *FEBS letters* 259.1 (1989): 153-157.

36. Cravioto MDC., *et al*. "A new inherited variant of the 3β-hydroxysteroid dehydrogenase-isomerase deficiency syndrome: evidence for the existence of two isoenzymes". *The Journal of Clinical Endocrinology and Metabolism* 63.2 (1986): 360-367.

37. Simard J., *et al*. "Molecular basis of congenital adrenal hyperplasia due to 3 beta-hydroxysteroid dehydrogenase deficiency". *Molecular Endocrinology* 7.5 (1993): 716-728.

38. Pletnev VZ., *et al*. "Rational Proteomics V: Structure-based mutagenesis has revealed key residues responsible for substrate recognition and catalysis by the dehydrogenase and isomerase activities in human 3β-hydroxysteroid dehydrogenase/isomerase type 1". *The Journal of Steroid Biochemistry and Molecular Biology* 101.1 (2006): 50-60.

39. Guruprasad K., *et al*. "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 Engineering, Design, and Selection* 4.2 (1990): 155-161.

40. Weerapana E and Imperiali B. "Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems". *Glycobiology* 16.6 (2006): 91R-101R.

41. Mitra N., *et al*. "N-linked oligosaccharides as outfitters for glycoprotein folding, form and function". *Trends in Biochemical Sciences* 31.3 (2006): 156-163.

42. Shimozawa O., *et al*. "Core glycosylation of cytochrome P-450 (arom) Evidence for localization of N terminus of microsomal cytochrome P-450 in the lumen". *Journal of Biological Chemistry* 268.28 (1993): 21399-21402.

43. Kaur J and Bose HS. "Passenger Protein Determines Translocation Versus Retention in the Endoplasmic Reticulum for Aromatase Expression". *Molecular Pharmacology* 85.2 (2014): 290-300.