

# In Silico Analysis, Structure Modeling and Phosphorylation Site Prediction of Vitellogenin Protein from *Gibelion Catla*

## Abstract

Vitellogenin is an egg yolk precursor protein expressed in the females of nearly all oviparous species including fish, amphibians, reptiles, birds, most invertebrates and monotremes. In the present study using bioinformatics tools and in silico modeling and analysis of Vitellogenin protein sequences of *Gibelion catla* was conducted. Primary structure prediction and physicochemical characterization were performed by computing theoretical isoelectric point (pI) 9.03, molecular weight 144862.03 Da, extinction co-efficient 78325, instability index 42.86 and aliphatic index 113.01. Grand Average of hydropathicity (GRAVY) was computed 0.212. Secondary structure assessment of Vitellogenin protein of carp *Gibelion catla* using GORIV reveals greater percentage of residues as alpha helix and random coils against the beta sheets. After performing homology modeling using Swiss model, a 3D structure of Vitellogenin of *Gibelion catla* have been predicted from its amino acid sequence. After the prediction structure has been validated through various validation tools. This homology modeling based structure will provide an insight to its functional aspects and further studies which are based on tertiary structure of protein.

**Keywords:** Structure modeling; Phosphorylation site; Exspasy peptide cutter; *Gibelion Catla*; Vitellogenin; EDCs; APOB

## Research Article

Volume 3 Issue 1 - 2017

**Rumpi Ghosh, Upadhayay AD and Roy AK\***

Bioinformatics Center, College of Fisheries, India

\*Corresponding author: AK Roy, Bioinformatics Center, College of Fisheries, CAU (I) Lembucherra, Tripura, India, Email: akroy1946@yahoo.co.in

Received: January 27, 2017 | Published: April 21, 2017

## Introduction

Vitellogenin (VTG or less popularly known as VG) is a precursor protein of egg yolk normally in the blood or hemolymph only of females that is used as a biomarker in vertebrates of exposure to environmental estrogens which stimulate elevated levels in males as well as females [1]. "Vitellogenin" is a synonymous term for the gene and the expressed protein. The protein product is classified as a glycolipoprotein, having properties of a sugar, fat and protein. It expressed in the females of nearly all oviparous species including fish, amphibians, reptiles, birds, most invertebrates and monotremes [2]. Vitellogenin is the precursor of the lipoproteins and phosphoproteins that make up most of the protein content of yolk. In the presence of estrogenic endocrine disruptive chemicals (EDCs), male fish can express the Vg gene in a dose dependent manner. Vg gene expression in male fish can be used as a molecular marker of exposure to estrogenic EDCs. Vitellogenin precursors are multi-domain Apo lipoproteins (proteins that bind to lipids to form lipoproteins), that are cleaved into distinct yolk proteins. At the N-terminal, it contains a sequence of about 670 amino acids (LV1n chain), which (MTP) and apolipoprotein B-100 (APOB). It is usually synthesized by the extra ovarian tissues of female animals, secreted into bloodstream and transported to ovary, where it is internalized by growing oocytes and proteolytically cleaved to form yolk proteins that are later used as the nutrients by developing embryos and larvae [3-8] However, Vg appears to have evolved pleiotropic functions in the advanced eusocial honeybee. It has been shown that honeybee

Vg is associated with quite a few biological processes including social organization, temporal division of labor and foraging specialization, regulation of hormonal dynamics and change in gustatory responsiveness [9-12]. Besides, honeybee Vg has been shown to be capable of reducing oxidative stress by scavenging free radicals, thereby prolonging lifespan in the facultatively sterile worker castes and reproductive queen castes [13]. Another novel function of Vg is linked with immune defense [14]. For example, Vg has recently been demonstrated to possess both hemagglutinating and antibacterial activities in the protochordate amphioxus (*Branchiostoma belcheri*) as well as the bony fish rosy barb (*Puntius conchonius*) [15,16]. SWISS-MODEL is a structural bioinformatics web-server dedicated to homology modeling of protein 3D structures. Homology modeling is currently the most accurate method to generate reliable 3-dimensional protein structure models and is routinely used in many practical applications. Homology (or comparative) modeling methods make use of experimental protein structures ("templates") to build models for evolutionary related proteins ("targets") [6].

The computed primary structure helps in quantitative study of protein-protein and protein-ligands interaction in solution. The aim to predict the secondary structure of protein based only on knowledge of their amino acid sequence. The secondary structure prediction displayed the alpha helix (H), beta strand (S), coil(C) and confidence score of prediction (0=low and 9=high for each residue). Structural class of protein can also be analyzed based on distribution of secondary structures elements. Long region of coil

elements in the protein usually indicate unstructured disordered region. The interactions that occur between the C, O and NH groups on amino acids in a polypeptide chain to form  $\alpha$ -helices,  $\beta$ -sheets, turns loops and other forms and that facilitate the folding into a 3-dimensional structure. 3D structure prediction By the Swiss model or other software is very easy within short time comparatively the output of experimentally determined protein structures typically by time-consuming and relatively expensive like X-ray crystallography or NMR spectroscopy is lagging far behind the output of protein sequences. The main aim of this study to analyze Vitellogenin protein and find utility of this protein in *Gibelion catla*.

## Methods

### Sequence analysis of vitellogenin protein of *Gibelion catla* (*catla catla*)

Vitellogenin protein of *Gibelion catla* was obtained from NCBI. The above obtained sequence was verified by Uniprot KB. The above obtained sequence was further used for complete protein sequence analysis (structural and functional annotation) and model building comparative modeling approach, Using ExPASy ProtParam Server complete primary structure analysis of protein has been performed, like molecular weight, theoretical isoelectric point (pI), amino acid composition, atomic composition, extinction co-efficient, half-life, aliphatic index, instability index and grand average of hydropathicity index (GRAVY). The ExPASy ProtParam tool is used to analyze protein sequences, structures and physicochemical properties of protein model [17].

### Secondary structure prediction

SOPMA and GORIV were used for secondary structure prediction of protein sequence of Vitellogenin [1]. The comparative analysis of Vitellogenin protein by SOPMA and GORIV is given in the Table 1. And the result is graphically represented in Figure 1 & 2.

**Table 1:** Percentage composition of all the secondary structure present in Vitellogenin of *Gibelion catla*.

Secondary Structure	SOPMA	GOR
Alpha Helix	51.76%(693)	55.27%(740)
310 Helix	0.00%	0.00%
Pi Helix	0.00%	0.00%
Beta Bridge	0.00%	0.00%
Extended strand	14.26%(191)	9.78%(131)
Beta Turn	7.92%(106)	0.00%
Bend Region	0.00%	0.00%
Random Coil	26.06%(349)	34.95%(468)
Ambiguous States	0.00%	0.00%
Other States	0.00%	0.00%
Sequence Length	1339	1339

### Homology modeling of vitellogenin protein

The modeling of the 3-dimensional structure of the protein was done using by homology modeling programs [3] Swiss model. Template was searched using Swiss model template library. A total 50 templates were found and it shows in Figure 3. The templates with highest quality have been selected for model building. Models are built based on the target template alignment model quality. Model quality was estimating assessing the QMEAN score [4]. QMEAN is a composite scoring function for both the estimation of the global quality of the entire model as well as for the local per-residue analysis of different regions within a model [18]. In the Swiss Model Workspace the QMEAN4 score is used to evaluate the generated models.

### ExpASy peptide cutter

The main aim of this work is to design a protein which is easily digestible in human gastrointestinal tract and thus in silico digestibility was performed using the ExpASy Peptide cutter tool. Peptide cutter is used to predict the cleavage [19].

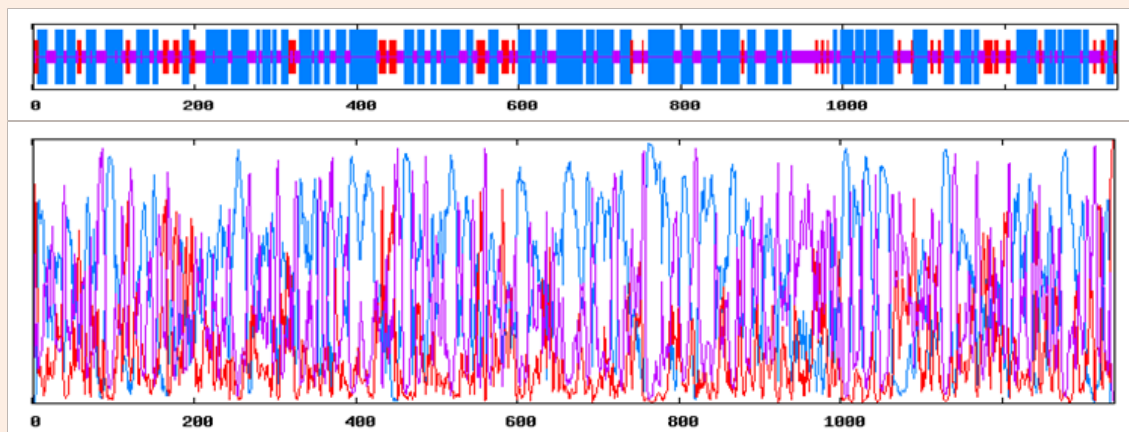
## Results and Discussion

### Primary structure prediction and physicochemical characterization of protein

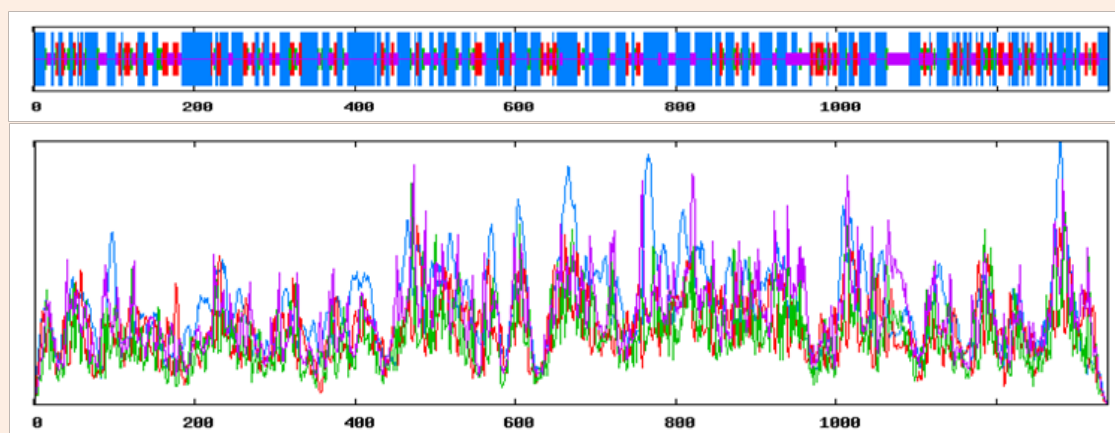
Using ProtParam tool provided by ExPASy primary structure have been performed it has been found that it is 1339 amino acid long residues of protein with 144862.03 Da molecular weight, whereas in Vitellogenin in *Bactericera cockerelli* (sulc) number of amino acid 2113 and molecular weight 234789.41 Da. In *Gibelion catla* theoretical isoelectric point or pI 9.03 which indicate its base property and Vitellogenin in *Bactericera cockerelli* (sulc) pI 9.18 which also base, whereas pI in Bovine  $\alpha$  S1 Casein showed 5.77 which indicate that the protein is negatively charged and it can be precipitate in acidic medium. At 280 nm its extinction co-efficient has been calculated by the tool and it is found to be 78325 with its unit, on the other hand, the extinction co-efficient of Vitellogenin in *Bactericera cockerelli* (sulc) is  $161650\text{M}^{-1}\text{cm}^{-1}$  and the aquaporin protein from catfish is  $36950\text{M}^{-1}\text{cm}^{-1}$ . The computed extinction co-efficient will provide help in quantitative study of protein-protein and protein-ligands interaction in solution [1]. Another important parameter called instability index has also been calculated and is computed to be 42.86. And this classified the protein is unstable. In stability index provides the stability of protein in laboratory condition or in test tube. If the value of instability index is less than 40, then it is considered as stable protein [2], on the other hand Vitellogenin in *Bactericera cockerelli* (sulc) instability index shows 48.45 which also unstable, whereas aquaporin protein in catfish calculated instability index found 30.22 which indicate that the protein is stable and in Bovine  $\alpha$  S1 Casein showed 56.03 which indicate also unstable. Aliphatic index of protein was found to be 113.01 which indicate this protein can survive at wide temperature range as the value is quite high, on the other hand in aquaporin protein showed 108.67 and Vitellogenin in *Bactericidal cockerelli* (sulc) 64.13. Grand Average of hydropathicity (GRAVY) was computed 0.212. It provides interaction with water of a particular protein. Low value of GRAVY index indicates better

interaction. Here it is 0.212 which indicates better interaction of Vitellogenin with water and Vitellogenin in *Bactericera cockerelli* (sulc) Gravy shows -0.904. On the other hand, 0.65 of aquaporin in catfish indicate better interaction with water. Estimated half-life of this protein is 30, whereas Vitellogenin of *Bactericera cockerelli* (sulc) also 30. The half-life of would be 30h in mammalian reticulocytes (in vitro), >20h in yeast (in vivo) and >10h in *E. coli*. The half life is the prediction of time taken by the protein to

reduce to half of its amount in the cell after its synthesis in the cell. Total no. of atoms is 20896 and formula  $C_{6549}H_{10704}N_{1756}O_{1848}S_{39}$ . Total number of negatively charged residues (Asp + Glu): 129, Total number of positively charged residues (Arg + Lys): 147. And maximum % of amino acid Alanine 14.7 % (197) and minimum % of amino acid tryptophan 0.4% (6), whereas in Vitellogenin of *Bactericera cockerelli* (sulc) negatively and positively charged residues are 211 and 247.



**Figure 1:** Secondary Structure Annotations of Vitellogenin from *Gibelion catla* using GORIV (h- Helix, e- extended strand, c-Random coil).



**Figure 2:** Secondary Structure Annotations of Vitellogenin from *Gibelion catla* using SOPMA (h- Helix, e- extended strand, c-Random coil).

### Secondary structure prediction

Using SOPMA and GORIV (tool for secondary structure prediction) regions of different secondary structure have been located in the entire protein sequence and it was found that the secondary structure is composed of alpha helix and beta sheets. Table 1 represents the comparative analysis of GORIV and SOPMA from which it is clear that alpha helix is predominantly present when the structure is predicted by both SOPMA and GORIV, followed by extended strand. After secondary structural annotation analysis, it has been found that highest percentage is of alpha helix (51.76%) and then random coil (26.06%), extended strand has been found to be (14.26%), beta turn (7.92%). And

other secondary structure like  $3_{10}$  helix, pi helix, Beta Bridge, bend region, ambiguous states and other states are totally absent. Higher percentage of helix indicates its helical chemistry (Figure 1, 2 & Table 1).

### Modeling of vitellogenin protein of Gibelion catla

3-dimensional structure of the protein (Figure 3) is predicted using the program called Swiss model using 1lsh.1 template from Swiss model template library for homology modeling. The evaluation of the predicted structure generated by Swiss model in Figure 4. Computational prediction of 3D structure of protein is based on Ab-initio, threading or Homology modeling. If the

sequence similarity search of target sequence is more than 60%, then homology modeling can be useful for 3D structure prediction of target protein. Homology modeling was done using a template sequence whose structure has been solved by either X-ray diffraction or NMR Spectroscopy. Swiss model is online tool for 3D structure prediction based on homology modeling [4,20]. This model quality is estimated by based on the Q mean score. Here Q means score Showed -3.70. 3D structure prediction By the Swiss model are very easy within short time comparatively the output of experimentally determined protein structures typically by time-consuming and relatively expensive like X-ray crystallography or NMR spectroscopy [21].

### ExpASy peptide cutter

For the proper metabolism, digestion and absorption of protein in humans, it has to be cleaved or hydrolyzed with digestive enzymes like Trypsin, pepsin, chymotrypsin, Thrombin, clostripain and lysine to yield peptides and free amino acids. The server shows the probability of the sequence cleaved by many enzymes. The total number of cleavages for chymotrypsin (high and low specificity), Arg-C protein's, were found to be 68 and 261 Asp-N endopeptidase 129, Glutamyl endopeptidase 92, Clostripain60, Formic acid 37 which indicate the digestion rate of the protein is good.

Name	Title	Coverage	Identity	Method	Oligo State	Ligands
1lsh.1.A	LIPOVITELLIN (LV-1N, LV-1C)	36.59	35.44	X-ray, 1.9Å	hetero-oligomer	28 x UPL <sup>□</sup> , 7 x PLD <sup>□</sup>
1lsh.1.A	LIPOVITELLIN (LV-1N, LV-1C)	36.59	35.44	X-ray, 1.9Å	hetero-oligomer	28 x UPL <sup>□</sup> , 7 x PLD <sup>□</sup>
3jco.1.8	26S proteasome regulatory subunit RPN1	15.10	15.10	EM	hetero-oligomer	None
4cr3.1.6	26S PROTEASOME REGULATORY SUBUNIT RPN1	15.10	15.10	EM, 7.7Å	hetero-oligomer	None
4cr3.1.6	26S PROTEASOME REGULATORY SUBUNIT RPN1	15.10	15.10	EM, 9.3Å	hetero-oligomer	None
4cr4.1.6	26S PROTEASOME REGULATORY SUBUNIT RPN1	15.10	15.10	EM, 8.8Å	hetero-oligomer	None
3jcp.1.8	26S proteasome regulatory subunit RPN1	15.10	15.10	EM	hetero-oligomer	None
5f0g.1.5	26S proteasome non-ATPase regulatory subunit 2	13.58	13.58	EM	hetero-oligomer	1 x ZN <sup>□</sup> , 6 x ATP <sup>□</sup>
5f0c.44.A	26S proteasome non-ATPase regulatory subunit 2	13.17	13.17	EM	monomer	None
5gqk.1.j	26S proteasome non-ATPase regulatory subunit 2	13.17	13.17	EM, 4.5Å	hetero-oligomer	6 x ADP <sup>□</sup>
5h4k.1.E	26S proteasome non-ATPase regulatory subunit 2	13.17	13.17	EM, 4.5Å	hetero-oligomer	None
4d4z.1.A	DEOXYHYPPUSINE HYDROXYLASE	14.22	14.22	X-ray, 1.7Å	monomer	1 x B3P <sup>□</sup> , 2 x FE <sup>□</sup> , 1 x OH <sup>□</sup>
4d50.1.A	DEOXYHYPPUSINE HYDROXYLASE	14.22	14.22	X-ray, 1.7Å	monomer	2 x FE <sup>□</sup> , 1 x PER <sup>□</sup> , 4 x GAI <sup>□</sup>
4d50.1.A	DEOXYHYPPUSINE HYDROXYLASE	18.81	18.81	X-ray, 1.7Å	monomer	2 x FE <sup>□</sup> , 1 x PER <sup>□</sup> , 4 x GAI <sup>□</sup>
4d4z.1.A	DEOXYHYPPUSINE HYDROXYLASE	18.81	18.81	X-ray, 1.7Å	monomer	1 x B3P <sup>□</sup> , 2 x FE <sup>□</sup> , 1 x OH <sup>□</sup>
4zv6.1.A	AlphaRep-7	17.76	17.76	X-ray, 2.2Å	hetero-oligomer	None
1lsh.1.B	LIPOVITELLIN (LV-2)	27.18	27.18	X-ray, 1.9Å	hetero-oligomer	28 x UPL <sup>□</sup> , 7 x PLD <sup>□</sup>
2xc3.1.B	ELONGATION FACTOR 3	13.82	13.82	X-ray, 2.7Å	homo-dimer	None
2xc8.1.A	ELONGATION FACTOR 3A	13.82	13.82	EM, 6.0Å	monomer	None
2wh1.1.A	ELONGATION FACTOR 3A	13.82	13.82	X-ray, 3.0Å	homo-dimer	2 x ANP <sup>□</sup>
2wh1.1.B	ELONGATION FACTOR 3A	13.82	13.82	X-ray, 3.0Å	homo-dimer	2 x ANP <sup>□</sup>

Figure 3: Swiss model Template library result.

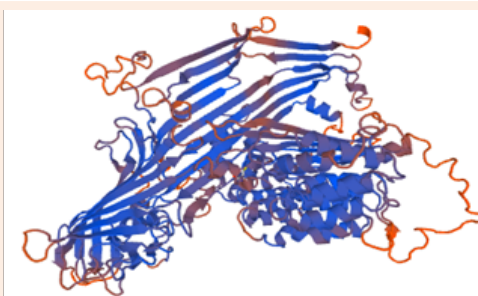


Figure 4: Model Structure of Vitellogenin protein of the *Gibelion Catla* 1lsh.1A (Lipovitellin as template).

### Phosphorylation site prediction

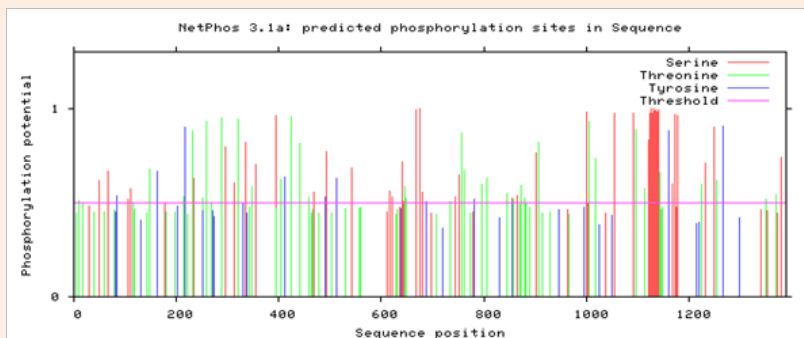
Phosphorylation site has been located in the NetPhosK 3.1L Server. Different kinases like CAM-GSK3, CKII, cdc2, CK1, DNAPK, p38MAPK, PKG, ATM, RSK, PKA, cdk5, PKB and UNSP have been found to be involved in the phosphorylation of protein. Highest score was predicted for the site 0.908 having serine residue. After analysis, % composition of amino acids has been obtained and is shown in Figure 5.

### Scanprosite result

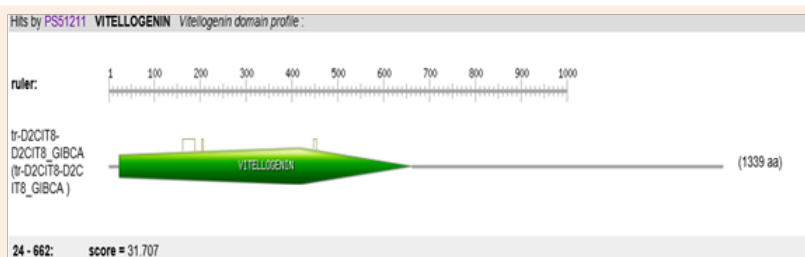
In Scanprosite result it showed that 1339 amino acid sequence are present and three disulphide bonds are present and their

position between 260-290, 200 and 450 and it represent by Figure 6. Scanprosite is a new and improved version of the web-based tool for detecting PROSITE signature matches in protein sequences [22]. Both user-defined sequences and sequences from the UniProt Knowledgebase can be matched against custom patterns or against PROSITE signatures [23]. To predict protein

function, assign family identity or detect remote homologues, searches against signature databases also known as secondary databases [1] are essential, Scanprosite provides a web interface to identify protein matches against signature from the prosite database [2].



**Figure 5:** Phosphorylation Plot obtained from NetPhosK3.1 Server.



**Figure 6:** Graphical representation of Scanprosite Result viewer.

## Conclusion

Vitellogenin protein is one of the most important proteins, in this study Vitellogenin protein of *Gibelion Catla* was selected and it's complete in silico analysis performed. Primary structure analysis and Physico chemical Characterization was performed by calculating various indices, GRAVY- 0.212 it is calculated by adding hydropathy values of each amino acid residues present in the protein sequence. The positive value indicated greater hydrophobicity and protein is sparingly soluble in water. The pI value can affect the solubility of a molecule at a given pH. Such molecules have minimum solubility in water or salt solutions at the pH that corresponds to their pI and often precipitate out of solution. Molecular weight 144826.03 Da and pI 9.03 which indicate that the protein is positively charged and that it can precipitate in base medium. Another important parameter called instability index has also been calculated and is computed to be 42.86. And this classifies that the protein Vitellogenin in *Gibelion catla* is unstable and easily able to degrade. Vg plays an integrative function in regulating immunity via its pleiotropic effects on both recognizing pathogen-associated molecular patterns and promoting macrophage phagocytosis. Our results suggest that fish Vg plays an integrative role in regulating innate immunity via its recognizing microbial cell wall constituents.

## Acknowledgment

Authors are thankful to the Dean, college of fisheries, Central Agriculture University, Lembucherra, Agartala for encouragement and support. The Financial assistant by DBT, GOI, New Delhi, India for BIF project under which this study has been carried out, is duly acknowledged.

## References

1. Singh A, Chaube R (2014) Bioinformatics analysis, structure modeling and active Site Prediction of Aquaporin protein from catfish *Heteropneustes fossilis*. *International Journal on Recent and Innovation Trends in Computing and Communication* 2(10): 2308-3215.
2. Guruprasad K, Reddy BV, Pandit MW (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* 4(2): 155-161.
3. Bandari M, Swaroop Kumar RS, Jagadeesh K, Kishnajuna Reddy CV (2013) In silico analysis and 3D structure prediction of Human Gastrin protein. *IJARIE* 2(4).
4. Aman Singh B, Roma L (2013) In silico sequence analysis of, structure prediction and Evaluation of fatty Acid Synthase

- Protein (FAS) of Gallus Gallus (Chicken) by Homology Modelling. International Journal of Science and Research 4(9): 843-847.
- Li Z, Zhang S, Liu Q (2008) Vitellogenin Functions as a Multivalent Pattern Recognition Receptor with an Opsonic Activity. PLoS One 3(4): e1940.
  - Schwede T, Kopp J, Guex N, Peitsch MC (2003). SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 31(13): 3381-3385.
  - Byrne BM, Gruber M, Ab G (1989) The evolution of egg yolk proteins. Prog Biophys Mol Biol 53(1): 33-69.
  - Dhadialla TS, Raikhel AS (1990) Biosynthesis of mosquito Vitellogenin. J Biol Chem 265(17): 9924-9933.
  - Amdam GV, Norberg K, Hagen A, Omholt SW (2003) Social exploitation of Vitellogenin. Proc Natl Acad Sci USA 100(4): 1799-1802.
  - Amdam GV, Norberg K, Page RE, Erber J, Scheiner R (2006) Down regulation of Vitellogenin gene activity increases the gustatory responsiveness of honey bee workers (*Apis mellifera*). Behav Brain Res 169(2): 201-205.
  - Guidugli KR, Nascimento AM, Amdam GV, Barchuk AR, Omholt S, et al. (2005) Vitellogenin regulates hormonal dynamics in the worker caste of a eusocial insect. FEBS Lett 579(22): 4961-4965.
  - Nelson CM, Ihle KE, Fondrk MK, Page RE, Amdam GV (2007) The gene Vitellogenin has multiple coordinating effects on social organization. PLoS Biol 5(3): e62.
  - Seehuus SC, Norberg K, Gimsa U, Krekling T, Amdam GV (2006) Reproductive protein protects functionally sterile honey bee workers from oxidative stress. Proc Natl Acad Sci U S A 103(4): 962-967.
  - Amdam GV, Hartfelder K, Norberg K, Hagen A, Omholt SW (2004) Altered physiology in worker honey bees (Hymenoptera: Apidae) infested by the mite *Varroa destructor* (Acari: Varroidae): a factor in colony loss during over-wintering? J Econ Entomol 97(3): 741-747.
  - Shi X, Zhang S, Pang Q (2006) Vitellogenin is a novel player in defense reactions. Fish Shellfish Immunol 20(5): 769-772.
  - Zhang S, Sun Y, Pang Q, Shi X (2005) Hemagglutinating and antibacterial activities of Vitellogenin. Fish Shellfish Immunol 19(1): 93-95.
  - Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, et al. (2005) Protein identification and analysis tools on the EXPASYserver. The Proteomics Protocols Handbook, USA, pp. 571-607.
  - Dhurigai N, Daniel RR, Auxilia LR (2014) Structure determination of Leghemoglobin using homology modeling. Int J Curr Microbiol App Sci 3(10): 177-187.
  - Appaiah P, Vasu P (2016) In silico Designing of Protein Rich in Large Neutral Amino Acids Using Bovine  $\alpha$ 1 Casein for Treatment of Phenylketonuria. J Proteomics Bioinform 9(11): 287-297.
  - Vidhya VG, Bhaskar A, Purushothaman P (2011) In silico analysis and 3D modeling of SORD Protein in Diabetic Retinopathy. J Comput Method Mol Design 1(4): 22-27.
  - Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereo chemical quality of protein structures. J Appl Cryst 26: 283-291.
  - Colovos C, Yeates TO (1993) Verification of protein structure: patterns of non-bonded atomic interactions. Protein Sci 2(9): 1511-1519.
  - De Castro E, Sigrist CJ, Gattiker A, Bulliard V, Genevaux PSL, et al. (2006) Scan Prosite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res 34: W362-W365.