Study of Vitellogenin Motif

Shashank Rana[1]

[1]Research Scholar,Vinayaka Missions University, Salem

Raghvendar Singh[2]

[2]Principal Scientist, National Research Centre on Camel, Bikaner

Shrikant Sharma[3]

[3]Researcher, Biochemistry Lab, National Research Centre on Camel, Bikaner

ABSTRACT

Vitellogenin (VTG) is known to be a female-specific protein and a precursor form of egg yolk. Threevitellogenin,Vtg I,Vtg II and Vtg III have been identified in chicken.Vitellinogen precursors supply the major egg yolk proteins which are a source of nutrients during early development of egg-laying (oviparous) vertebrates and invertebrates. A complete structural analysis and 3-D modelling of Vitellogenin protein of Chicken (Target), has been already done. The 1852 amino acid sequence of the Vitellogenin protein (Gallus gallus) was retrieved from NCBI database. OOPS model of Multiple EM for Motif Elicitation (MEME) online software package was used to carry out this analysis. Total fifteen amino acid sequences were taken for motif analysis.According to our results, Vitellogenin sequences, contains three motifs. We analyze that all fifteen sequences includes all threemotifs but their start points are different. Motif investigationsuggeststhat all selected VTG sequences from different resources have common conserved patterns.

KEYWORDS

MEME, Sequence Analysis, OOPS,Vitellogenin.

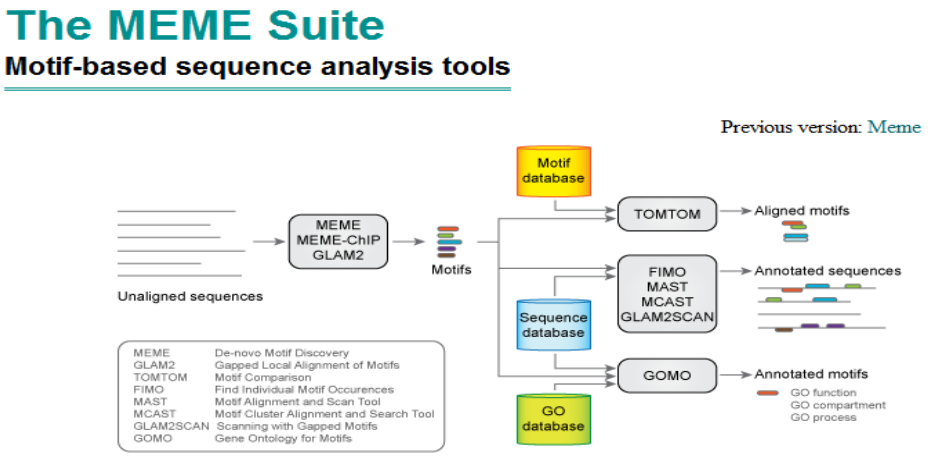
INTRODUCTION

Vitellogenin or VTG or VG(taken from Latinvitelluswhich means yolk and gener means to produce) is anidentical term for the gene and the uttered protein. VTG belongs to a family of some lipid transport proteins. Vitellogenin (VTG) is known to be a female-specific protein and a precursor form of egg yolk [1]. VTG precursors offer the major egg yolk proteins that are a good source of nutrients during early stage development of oviparous, vertebrates and invertebrates. VTG precursors are the multi-domain apolipoproteins, which cleaved into different yolk proteins. In vertebrates, aentirevitellinogen protein is composed of an N-terminal signal peptide (for export)and four regions (that can be cleaved into yolk proteins lipovitellin-1 (LVI), phosvitin (PV), lipovitellin-2(VII)), and a von-Willebrand factor type D domain (YGP40) [2,3].VTG protein synthesis take place in the liver under the influence of estrogen [4] and included into the developing oocytes through blood circulation.Therefore, Vitellogenin has been used as apointer for maturation of females [5,6,7] and for the early stage sex determination of several teleost fishes [8].

Vitellogenin protein, serve as a trace mineral transporting protein which actively mediate the transfer of these essential nutrients from liverof the hen to the ovary and developing oocyte, and therefore, to the yolk of the egg.Chicken Vitellogeninhas been considered to be proteolyticallycleaved into heavy and light chains namely lipovitellins and phosvitin respectively.These proteins are the major yolk granule proteins, during or after transportation into oocyte. Chicken α- and β-lipovitellin are derived from parent VTG proteins and have four subunits (125, 80, 40, and 30 kDa) and two subunits (125 and 30 kDa), respectively. In chicken, there arethree vitellogenins,namely.Vtg I,Vtg II and Vtg III have been characterized[9].Amino acid analyses indicated the presence of a highly phosphorylated serine-rich phosvitin in VtgI and vtgII (116 mol of Peach)and of a low molecular weight phosphopeptide (phosvette) in VTGIII (44 mol of P).According to Wallace and Morgan [10], unfractionated phosvitin is composed of 5 major components known as B (MW 40000D), C (MW 33000D), E (MW 115000D) E (MW 218000D) and F (MW 13000D).According to the stoichiometric considerations it is suggested that vitellogenin I gives rise to phosvitins C and F; vitellogenin II gives rise to phosvitin B, whereas vitellogenin III gives rise to either phosvitin E1 or E2, but not E1 & E2. Another forth one but yet undetected, Vitellogenin may also exist for the chicken [11].Taborsky and Mok [12] firstly determined the molecular weights of minor and major phosvitin as 36KDa and 40KDa. Egg yolk phospho-proteinis composed of phosvitin and phosvette.Phosvitin accounts for almost 60% of total egg yolk phosphoprotein and grasps about 90% of the egg yolk phosphorous. The meaning of phosvitinsuggests that it carried high amount of phosphorous and its origin of egg yolk.

Analysis of motifs present in vitellogenin protein has been done by using MEME (Multiple EM for Motif Elicitation) analysis which is present on http://meme.nbcr.net/meme4_6_1/ (Fig. 1).

Figure-1: MEME overview



Materials and Methods

We have selectednineteen amino acids sequences of Vitellogenin proteinfrom NCBIretrieved in a fasta format.[13].OOPS model of MEME (Multiple Expectation Maximization for Motif

Elicitation) has been used for analysis of motifs present in given sequence. In MEME, output demonstratescolor graphical alignment along with common regular expression of all possible motifs.The block represents start and end point of the sequences and demonstrates AA length.E-value explains thestatistical significance of the motif. MEME(ver.3.5.7)[14]usually predict the most statistically significant (low *E*-value) motifs first. As per programming of MEME suite,it is designed for prediction of up to three motifs and they may be present in some or all of the input sequences. MEME decides the width and number of occurrence of each motif mechanicallyon the bases of low E-value of the motif-the likelihood of finding auniformly well-conserved pattern in random sequences. By default software programming, only motif widths between six and fifty were considered, and MEME output is HTML and reveals the motifs as local multiple alignments of (subsets of) the given sequences, as well as in several other formats.Block diagrams of MEME suite, defines the relative positions of the motifs in each of the input sequences whereas, *E*-valuedescribes an estimate of the expected number of motifs with the given log likelihood ratio (or higher) and with the same width and site count, that one would find in a likewise sized set of random sequences.Further, width describes that; each motif explains a specific pattern of a fixed sequence in which no gaps are allowed.In MEME package,sitesdefine the conserved region present in the motifs.The number of sites is contributing to the construction of the motifs.In **MEME** suite, consequences also exhibit***information content*** of the motif in bits, It is defines to the amount of the ***uncorrected*** information content, R(), in the columns of the LOGO. Motifs are characterized by position-specific probability matrices which arespecifying the probability of each possible letter appeared at each probable position in an occurrence of the pattern (Motif). These are showed as "sequence LOGOS"includes stacks of letters at every position in the motif. The total height of the stack is known as the "**information content"** of that particular position in the motif (in bits). In case of proteins, the categories are based on the bio-chemical properties of the diverse amino acids.

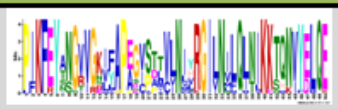
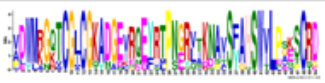
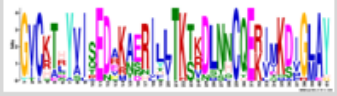
Results and Discussion

In MEME suite, results display in the form of graph whereas sequences will exhibits in the form of sequence logo/regular expression (Table-2). Motif overviewin figure-2 has shown 4.6e-649E-value of motif one, 8.3e-520E-value of motif two and 1.5e-494E-value of motif three.In Results were analyzed on the bases of e-value and p-value. Where second figure, describe about the conserved pattern of motifs and first one describe about the width of the same match. Higher p-value described the best match whereas lower the e-value better the results.Bysubmitting multiple amino acid sequences to MEME suite, we find that all the sequences have everyone three motifs but the starts points of all these motifs are vary sequence to sequence**.** Figureno 3, 5 and 7defines the number of site which contributing to the construction of the motif. In each protein sequence define the site in color format.These are shown aligned with each other. Each site is documented by the name of the sequence where it occurs, the strand, and the Start position in the sequence where the site begins. The sites are listed in increasing order based on statistical significance known as *p*-value.Figure4, 6 and 8 shows that block diagram of each site and represent the motif location.The frequency of the motif in the protein sequences are exposed as colored blocks on a line. Motif one was present in seq. 9, 10, 7, 8, 3, 13, 12, 11, 6, 17, 15, 18 and 4. On the other hand, motif two is identified in seq.no 1, 2, 3,4,6, 7,8, 9, 10, 11, 12, 13, 15, 16, 17,18 and 19 .On the bases of result analysis, Motif three present in seq.1, 2, 3,4, 5,6, 7, 8, 9, 10, 11,12,13, 15, 16, 17, 18 and19.The occurrence of motifshas been displayed in combined block diagram (Figure-9) and might not be closely comparableas reported in each motif segment.

Table-1: Scientific name of different species

Seq. No.	Scientific Name
1	***Homo sapiens***

2	***Pan troglodytes***
3	***Macacacyclopis***
4	***Oryctolaguscuniculus***
5	***Camelus dromedarius***
6	***Equuscaballus***
7	***Musmusculus***
8	***Susscrofa***
9	***Ovisaries***
10	***Capra hircus***
11	***Bostaurus***
12	***Bosindicus x Bostaurus***
13	***Bosindicus***
14	***Bosgrunniens***
15	***Rattusnorvegicus***

Sr. No.	Width/ Sites	E value	Logo	Regular expression
1	Width 50 Sites 19	4.6e-649		P[IF]KFEY[AT][NS]G[VR]VGK [IVL] [FY]AP[EA]G[VIC][SP] [TAD][TLM][VC]LNI[YHV]RGI [LV]N[IV]LQ[LI][NT]IKK[TS] QN[VI]Y[ED]LQE
2	Width 50 Sites 16	8.3e-520		[VA]DWM[KR]G[QK]TCG[LI] CGKADGE[VI]RQE[YF]RTPN[GE] R[VL]TKNAVS[FY]AHSW[VI]LP [AGS][EK]SCRD
3	Width 41 Sites 19	1.5e-494		G[VI]C[KHR]T[HLR]Y[VA]I[SQ] ED[AER]K[AN][ES]RIL[LV]TK[TS] [KRV]DLN[NH]CQE[KR][IV]MK [DS][IV]G[LM]AY

Name	Start	*p*-value	Sites		
sequence9	105	1.22e-58	TSALAAQLLT	PIKFEYANGVVGKVFAPAGVSATVLNIYRGILNILQLNIKKTQNVYELQE	PGAQGVCKTH
sequence10	105	1.38e-56	TSALAAQLLT	PIKFEYANGVVGKLFAPAGISTTVLNILRGIVNILQLNIKKTQNVYELQE	PGVQGICKTH
sequence7	105	1.38e-56	TSALAAQLLT	PIKFEYANGVVGKLFAPAGISTTVLNILRGIVNILQLNIKKTQNVYELQE	PGVQGICKTH
sequence8	105	2.67e-56	TSALAAQLMT	PIKFEYTNGVVGKMFAPEGVSTMVLNVYRGILNVLQLNIKKTHNVYELQE	AGAQGVCKTL
sequence3	105	1.80e-55	TQALAAQLLI	PIKFEYANGVVGRVFAPEGVSATVLNVHRGILNILQLNIKKNQNVYELQE	AGAQGTCKTD
sequence12	105	4.76e-55	TSALAAQLQI	PIKFEYANGVVGKVFAPAGVSPTVLNLHRGILNILQLNLKKTQNIYELQE	AGAQGVCKTH
sequence13	105	6.82e-55	TSALAAQLLT	PIKFEYINGVVGKIMAPEGISTLVLNIQRGILNVLQLNIKKTQNVYELQE	AGAQGVCKTL
sequence11	105	3.92e-54	MSTMTAELQI	PIKFEYTNGVVGKIFAPAGVSPIVLNFYRGILNIFQLNLKKTQNIYELQE	AGVQGVCKTH
sequence6	105	6.18e-54	TSALAAQLLT	PIKFEYANGVVGKVFAPAGISATVLNILKGIVNILQLNIKKTQNVYELQE	PGVQGICKTH
sequence17	105	1.47e-52	TSALSAQLLT	PIKFEYANGVIGKVYAPAGVSTAVLNIFRGILNIFQINIKKTMNVYDLQE	TGVKGVCKTS
sequence15	105	3.93e-51	TTALATQLST	PIKFEYTNGVVGRLAAPPGVSTTVLNIYRGIINLLQLNVKKTQNVYEMQE	SGAHGVCKTN
sequence18	98	8.70e-51	KDALAPQLVL	PIKFEYTNGIVGKLYAPEQIATLVLNIHRGILNILQLNIKKTHKVYDLQE	VGTQGVCKTL
sequence4	105	9.60e-51	TEALAAQLLI	PIKFEYANGVVG	

One

One

ame	Start	*p*-value	Sites		
sequence6	159	1.21e-49	VYELQEPGVQ	GICKTHYVISEDVKAERILLTKTKDLNNCQERIMKDIGLAV	TEKCVECEAR
sequence10	159	4.21e-48	VYELQEPGVQ	GICKTHYVISEDVMAERILLTKTKDLNNCQERIMKDIGLAV	TEKCVECESR
sequence7	159	4.21e-48	VYELQEPGVQ	GICKTHYVISEDVMAERILLTKTKDLNNCQERIMKDIGLAV	TEKCVECESR
sequence9	159	6.81e-48	VYELQEPGAQ	GVCKTHYVISEDAKADRILLTKTKDLNHCQERIVKDIGLAV	TERCVECEAR
sequence13	159	1.79e-46	VYELQEAGAQ	GVCKTLYAITEDEKAERILLTKSRDLNNCQEKIMKDIGLAV	TEKCIKCQQN
sequence8	159	1.98e-44	VYELQEAGAQ	GVCKTLYAIAEDEKAERILLTKTRDLNHCQEKIMKDLGLAY	TEKCAKCQQD
sequence3	159	5.22e-44	VYELQEAGAQ	GTCKTDYVISEDAKAERIYITKSKDLNNCQERIMQDIGMAV	TETCAHCQQN
sequence1	15	1.06e	VYDLQEVGT	GVCKTLYSISEDARIENILLTKTRDMNNCQERIMKDMG	TEKCDKCQR

ame	Start	p-value	Sites		
8	2	-41	Q	LAY	E
sequencel 5	15 9	4.22e -41	VYEMQESGA H	GVCKTNYVIREDARAERIHLTKTKDLNHCQEKIMKAIG LEH	VEKCHDCEA R
sequencel 6	15 9	5.77e -41	VFDLQEVGT Q	GVCKTLYSISEDARNENILLTKTRDLNNCQERLIKDMGL AY	TQKCERCQE E
sequencel 9	15 8	1.07e -40	IYDLQEEGV Q	GVCRTQYAITENEKTERILLTKSRNLNQCQEKVMKDIG LAY	TKTSHKYQQ D
sequencel	15 9	1.99e -40	VYELQEAGI G	GICHARYVIQEDRKNSRIYVTRTVDLNNCQEKVQKSIG MAY	IYPCPVDVM K
sequencel 1	15 9	8.15e -40	IYELQEAGV Q	GVCRTHYVISEDSKANHITVTKSKDLSHCQERITKDFGL AY	TEKCDECTE R
sequence2	15 9	1.48e -39	VYELQEAGI G	GVCHTRYVIQEDRKNSRVSVTKTVDQNNCQEKVIKSVG MAY	IYPCPVDMM K
sequencel 2	15 9	2.19e -39	IYELQEAGA Q	GVCRTHYVISEDPKANHITVTKSKDLSHCQERIVKDVRL AY	TERCAECTE R
sequencel 4	80	1.25e -38	VYELQEAGI G	GVCHARYVIQEDRKNSRIFVTKTVDSTNCQEKVEKSVG MAY	IYPCPVDMM K
sequence5	80	1.25e -38	VYELQEAGI G	GVCHARYVIQEDRKNSRIFVTKTVDSTNCQEKVEKSVG MAY	IYPCPVDMM K
sequence4	16 5	1.97e -36	AGAQGICKT D	YVISEDAVISEDAKAERIYITKSKDLNNCHERIMKDIGMA Y	TETCAQSQQ R
sequencel 7	15 9	9.95e -34	VYDLQETGV K	GVCKTSYVLKEDPKADRLHLTKTTDLNLCTEKINMDVG MAG	YTNKCEECV R

three

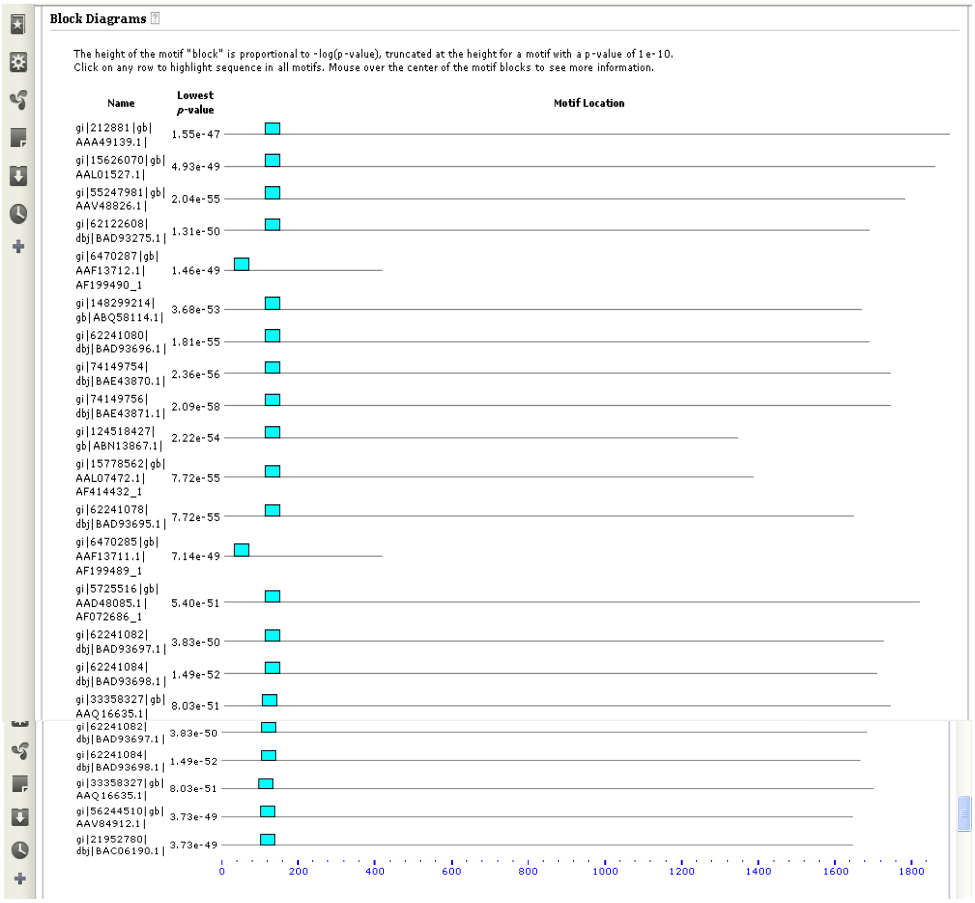


Figure 4.15Block Diagram ofVitellogenin Proteins Motif 1

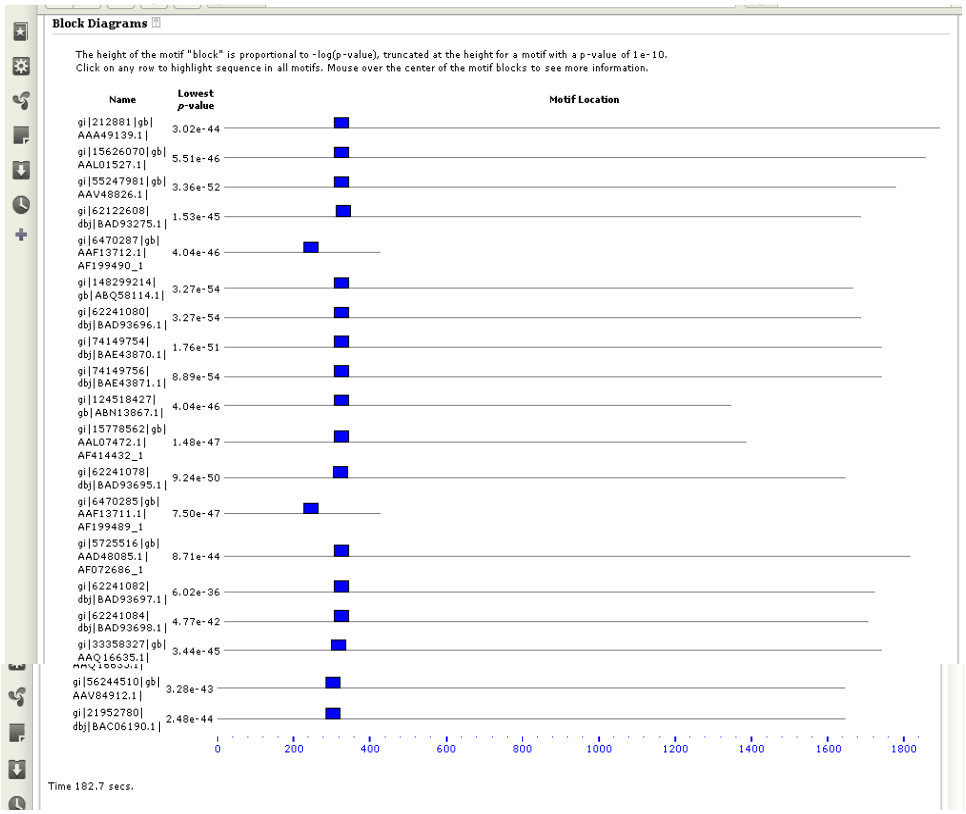


Figure 4.16Block Diagram Of Vitellogenin Proteins Motif 2

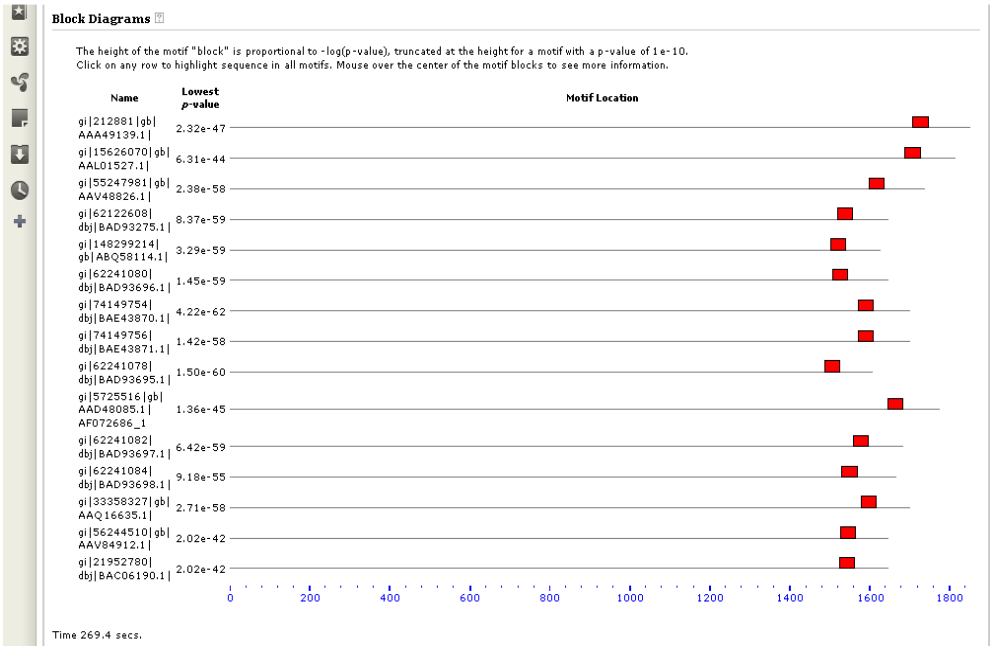
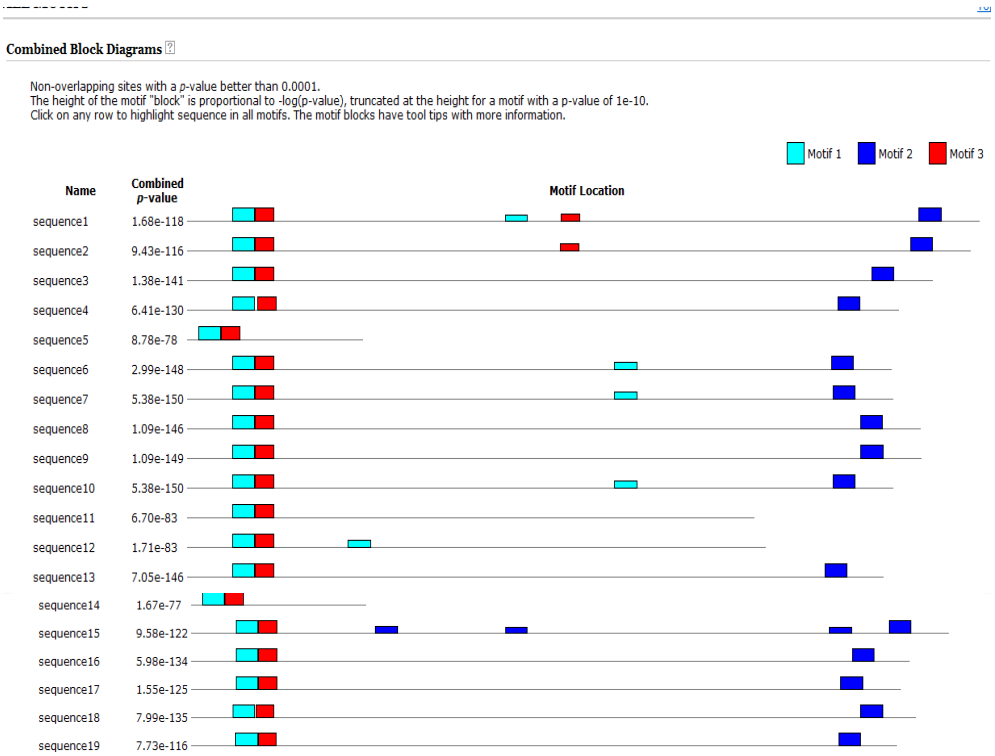


Figure 4.17Block Diagram Of Vitellogenin Proteins Motif 3



Conclusion

MEME suites developers provide an easy accessible tool for Motif analysis to the researcher community, who are interested in analyzing sequences of nucleic acids and proteins.At last,Our research explain that using multiple motifs gives much better database search results than using single motifs. Multiple motifs have more information about the quality of the protein family than do single motifs. On the bases of results, the presentstudy suggests that Vitellogeninsequence of different species has same conserve region or patterns. On the base of our result we assumed that vitellogeninshow similar function in all species. As per analysis of results, it describes that although vitellogenin presentsin different source of origin,they contains common patterns of amino acids. It is again well-knownfact that motifs may overlap with one another due to the reason of common consensus patterns.

References

1. Wallace, R.A., and K., Selman., 1981."Cellular and dynamics aspects of oocyte growth in teleosts".American journal of Zoology, 21: 325-343.
2. Finn, R.N., 2007. "Vertebrate Yolk Complexes and the Functional Implications of Phosvitins and Other Subdomains in Vitellogenins".Biology of Reproduction, 76 (6):926-935 –PMID 17314313.
3. Banaszak L.J., andThompson, J.R., 2002. "Lipid-protein interactions in lipovitellin".Biochemistry ,41 (30): 9398–9409-PMID 12135361.
4. Hara, A., T., Matsubara, M., Saneyoshi, and K., Takano, 1984. "Vitellogenin and its derivatives in egg yolk proteins of white spotted char (Salvelinusleucomaenis)". Bull. Fac. Fish. Hokkaido University, 35: 144-153.
5. Takemura, A., A., Hara, and K., Takano, 1991. " Immunochemical identification and partial characterization of female-specific serum proteins in white-edged rockfish, Sebastestaczanowskii". Environmental Biology of Fish, 30: 49-56.
6. Tao, Y., A., Hara, R., G., Hodson, L., C., Woods III, and C.,V., Sullivan, 1993. "Purification, characterization and immunoassay of striped bass (Moronesaxatilis) vitellogenin.Fish Physiology of Biochemistry,12: 31-46.
7. Matsubara, T. T., Wada, and A., Hara, 1994. "Purification and establishment of ELISA for vitellogenin of Japanese sardine (Sardinopsmelanostictus)". Computational Biochemistry of Physiology, 109: 545-555.
8. Takemura, A., M., Oka, and M., Kanematsu,1996. "Early sex distinction in greater amberjack, Serioladumerili, using skin mucus". Bulletin of the Japanese Society for the Science of Fish, 62: 62-67.
9. Wang, S.Y., and Williams, D.L., 1980. "Identificiation, purification, and characterization of two distinct avian vitellogenins". Biochemistry,19(8):1557-63.
10. Wallace,R.A., and Morgan J.P.,1986. "Chromatographic resolution of chicken phosvitin multiple macromolecular species in a classic vitellogenin-derived phosphoprotein".Biochemical Journal,240(3):871-8.
11. Robin, A. Wallace and Jay P., and Morgan, 1986. "Chromatographic resolution of chicken phosvitinMultiple macromolecular species in a classic viteliogenin-derived phosphoprotein".Biochemical Journal, 240(3):871-878.
12. Taborsky, G., andMok, C.C., 1967."Phosvitin.Homogeneity and molecular weight".Journal of Biological chemistry, 242 (7):1495-1501.
13. http://www.ncbi.nlm.nih.gov/
14. http://meme.nbcr.net/meme4_6_1/cgi-bin/meme.cgi