

IDENTIFICATION OF MOTIFS IN BIOACTIVE PEPTIDES PRECURSORS**Shrikant Sharma**^{*1}, Raghvendar Singh², Shashank Rana¹, Ravi Uniyal³¹*School of Life Science, Singhania University, Jhunjhunu, (RJ)²Bioinformatics Facility, Department of Immunology, College of Biotechnology, Sardar Vallabh Bhai Patel University of Agriculture & Technology, Meerut (U.P.)³. LLRM Medical College, Meerut (U.P.)**E-mail of Corresponding Author:** shribioinfo@gmail.com**Abstract**

In-silico study was carried out to study motifs present in precursors of antimicrobial, antithrombotic, casein derived and mineral binding bioactive peptides by using online servers. MEME suite was used for defining the common consensus pattern present in bioactive peptides precursors. It was found that although three different consensus patterns was identified in precursors but these consensus pattern overlapped in many sequences and all three motifs may be present in many peptides precursor sequences.

Keywords: *In-silico* study; bioactive peptides; MEME Suite; Consensus pattern.**1. Introduction**

Bioactive peptides have been defined as specific protein fragments that have a positive impact on body functions and conditions and may ultimately influence health¹. According to Fitzgerald & Murray (2006)⁴ bioactive peptides have been defined as 'peptides with hormone- or drug- like activity that eventually modulate physiological function through binding interactions to specific receptors on target cells leading to induction of physiological responses. Most of the bioactivities of bioactive peptides are encrypted within the primary sequence of the native protein and peptides require to be released through one of the following ways: Hydrolysis by digestive enzymes such as trypsin and pepsin^{5,7}, Food processing⁶ and Through hydrolysis by proteolytic microorganisms or through the action of proteolytic enzymes derived from the microorganisms⁴. MEME (Multiple EM for Motif Elicitation) is one of the most widely used tools for searching for novel 'Signals' in sets of biological sequences. Applications include the discovery of new transcription factor binding sites and protein domains. MEME is an unsupervised learning algorithm for discovering motifs in sets of protein or DNA sequences. To use MEME via the website, the user provides a set of sequences in the FASTA format by either uploading a file or by cut-and-paste. The only other required input is an email address where the results will be sent. The purpose of MEME (Multiple EM for Motif Elicitation) (rhymes with 'team') (1,

2) is to allow users to discover signals (called 'motifs') in DNA or protein sequences. Functional similarities between milk and blood coagulation as well as sequence homologies exist in the fibrinogen g-chain and k-casein⁹. Jolles *et al.*, (1986)⁸ showed that bovine k-casein f106-116 inhibited platelet aggregation and combined with the receptor site, consequently preventing fibrinogen binding with blood platelets. This inhibition was dependent on peptide concentration. The two smaller tryptic peptides (k-casein f106-112 and f113-116) exerted a much more minimal effect on platelet aggregation and did not inhibit fibrinogen binding. These peptides are referred to as casoplatelins. The behaviour of k-casein f106-116 is similar to that of the C-terminal peptide of the human fibrinogen g-chain (Fiat *et al.*, 1989). The mechanism involved in milk clotting, defined by interaction of kappa casein with chymosin bear a remarkable similarity to the process involved in blood clotting, defined by interaction of fibrinogen with thrombin¹⁰. Structural homologies between cow κ -casein and human fibrinogen γ -chain were found by Jolle's and Caen⁹. The κ -casein fragment named casoplatelins, obtained from tryptic hydrolysates, shows antithrombotic activity by inhibiting fibrinogen binding platelets^{9,10}. These peptides are released during gastrointestinal digestion and absorbed intact into the blood, which supports the concept that they exert an antithrombotic effect in vivo. The potential physiological effects of these antithrombotic peptides have not been

established, but such peptides have been detected in the plasma of newborn children after breastfeeding or ingestion of cow milk-based infant formulae¹¹.

2. Materials and Methods

We have select nineteen amino acids sequences of antithrombotic bioactive peptides in fasta format retrieved from NCBI. We use OOPS model of MEME (Multiple EM for Motif Elicitation) for finding out motifs present in given sequence. The output shows colour graphical alignment and it also displays common regular expression of motifs and the block represent start and end point of the seq and shows AA length. E-value describes the statistical significance of the motif. MEME usually finds the most statistically significant (low E-value) motifs first. The E-value is 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 similarly sized set of random sequences. On the other hand width describes that each motif describes a pattern of a fixed width as no gaps are allowed in MEME motifs. Sites in the MEME package define the conserved region present in the motifs. The number of sites contributes to the construction of the motif. The **information content** of the motif in bits, It is equal to the sum of the **uncorrected** information content, $R()$, in the columns of the LOGO. This is equal relative entropy of the motif relative to a uniform background frequency model. Whereas relative entropy of the motif, computed in bits and relative to the background letter frequencies given in the command line summary. It is equal to the log-likelihood ratio (llr) divided by the number of contributing sites of the motif times $1/\ln(2)$, $re = llr / (\text{sites} * \ln(2))$.

Sequence LOGO

MEME motifs are represented by position-specific probability matrices that specify the probability of each possible letter appearing at each possible position in an occurrence of the motif. These are displayed as "sequence LOGOS", containing stacks of letters at each position in the motif. The total height of the stack is the "information content" of that position in the motif in bits. For proteins, the

categories are based on the biochemical properties of the various amino acids.

3. Results and Discussion

By default, MEME looks for up to three motifs, each of which may be present in some or all of the input sequences. MEME chooses the width and number of occurrences of each motif automatically in order to minimize the 'E-value' of the motif—the probability of finding an equally well-conserved pattern in random sequences. By default, only motif widths between 6 and 50 are considered. The MEME output is HTML and shows the motifs as local multiple alignments of (subsets of) the input sequences, as well as in several other formats. 'Block diagrams' show the relative positions of the motifs in each of the input sequences. After the submission of sequences in query box of MEME, results display in the form of graph and seq will display in the form of sequence logo or regular expression. Here sequences are given as seq1, seq2 etc. and their accession numbers are given in table 1.9. Due to restriction of pages, figure and tables of motif one discovered in each peptides type are given here and along with it details of peptides sequence taken for study are also given. On the submission of multiple amino acid sequences, Antithrombotic peptides, we find that there are three motifs present in these sequences. Motif one is present in seq 6, 8, 9, 10, 11, 12, and 13. On the other hand motif two is found in seq 6, 9, 10, 11, 12 and 13. It is again noted that sequence no 6, 8, 9, 10, 11, 12, 13 and 18 carry motif three. On the basis of this analysis, we found that seq no. 6, 8, 9, 10, 11, 12 and 13 contains all three motifs but seq no. 1-5, 7, 14-17 and 19 contains no consensus region and seq 18 has consensus region like 6, 8-13.

We have taken thirty five opioid peptide precursor sequences and on the basis of analysis, we found that the consensus region belongs to Motif one universally present in opioid peptides precursors named seq no 32, 27, 26, 34, 14, 31, 29, 18, 19, 35 and 33. On the other hand consensus region belong to Motif two present in opioid peptides precursor seq no 34, 32, 14, 31, 29, 18, 35, 30, 1, 19, 17 and 33. Again we found that consensus pattern present in motif three belongs to the opioid peptides precursor seq no 32, 27, 26, 34, 14 and 30. We found that the motif one, two and

three present in seq no 34, 32, 14 and 30. Here motif presents in previously stated sequences may be overlaps.

We have taken ten mineral binding peptide precursor sequences and on the bases of analysis, we found that the consensus region belongs to Motif one universally present in mineral binding peptides precursors named seq no 2, 5 and 7. On the other hand consensus region belong to Motif two present in opioid peptides precursor seq no 6 and 7. Again we found that consensus pattern present in motif three belongs to the opioid peptides precursor seq no 5 and 6. We found that the motif one and two three present in seq no 7. But the motif one and three share their sequence with se no 6. Seq no 5 carry consensus pattern of motif one and three.

We have taken eleven immunomodulatory peptide precursor sequences and on the bases of analysis, we found that the consensus region belongs to Motif one universally present in immunomodulatory peptides precursors named seq no 3 and 4. On the other hand consensus region belong to Motif two present in immunomodulatory peptides precursor seq no 2 and 9. Again we found that consensus pattern present in motif three belongs to the opioid peptides precursor seq no 1 and 5.

In case of combined study of different peptides, we have taken one hundred and

eleven peptide precursor sequences of antimicrobial, immunomodulatory, opioid and casein peptide precursor and on the bases of analysis, we found that the consensus region belongs to Motif one universally present in peptides precursors named seq no 10, 6, 11, 12, 8, 9 and 15. On the other hand consensus region belong to Motif two present in peptides precursor seq no 12, 10, 6, 11, 8, 9 and 13. Again we found that consensus pattern present in motif three belongs to the precursor seq no 12, 10, 6, 11, 8, 9 and 13. We found that the motif one, two and three present in seq no 10, 12, 11, 6, 8 and 9. Interestingly motif two and three also share their sequence with seq no 13.

In case of casein related peptides, we have taken thirty seven casein peptide precursors and on the bases of analysis, we found that the consensus region belongs to Motif one universally presents in peptides precursors named seq no 32, 27, 26, 34, 14, 31, 29, 18, 19, 35 and 33. On the other hand consensus region belong to Motif two present in opioid peptides precursor seq no 1, 14, 17, 18, 19, 29, 30, 31, 32, 33, 34 and 35. Again we found that consensus pattern present in motif three belongs to the precursor seq no 12, 10, 6, 11, 8, 9 and 13. We found that the motif one, two and three present in seq no 32, 26, 27, 34, 14 and 30.

Figure1.1: Conserved pattern of Antithrombotic peptide

DISCOVERED MOTIFS



Figure 1.2: MEME output. Block diagram of motif present in antithrombotic peptide This portion of an MEME HTML output form shows a protein motif that MEME has discovered in the input sequences.

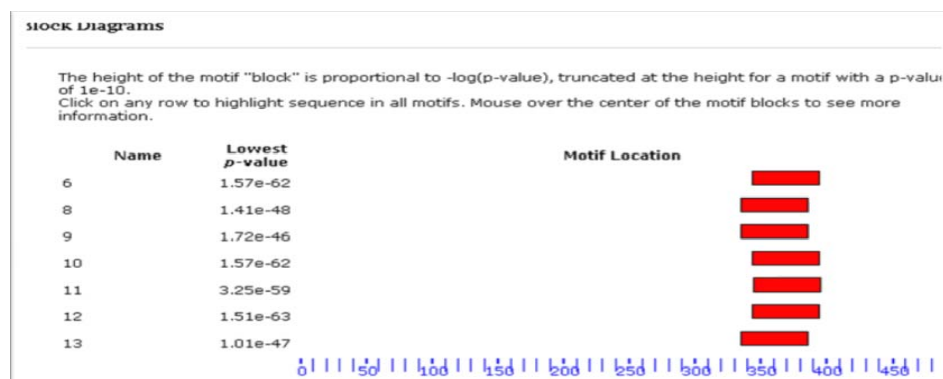


Table1.1: The sites identified as belonging to the motif are indicated, and above them is the 'consensus' of the motif and a color-coded bar graph showing the conservation of each position in the motif.

Name	Start	p-value	Sites		
12	344	1.51e-63	PFTESACL LL	IQPHYASDLKVEGLTFQQNSLNWMKKLSPRAIHLTMPQLVLQGSYD LQD	LLAQAEIPAI
10	344	1.57e-62	PFTESACL LL	IQPHYASDLKVEGLTFQQNSLNWMKKLSPRTHLTMPQLVLQGSYD LQD	LLAQAEIPAI
6	344	1.57e-62	PFTESACL LL	IQPHYASDLKVEGLTFQQNSLNWMKKLSPRTHLTMPQLVLQGSYD LQD	LLAQAEIPAI
11	345	3.25e-59	PFTESACL LL	IQPHYASDLKVEGLTFQQNSLNWMKKLSPRAIHLTMPRLVLRGSYD LQD	LLAQAEIPAI
8	335	1.41e-48	PLGESVTL LL	IQQCASDLDRVEVLVFQHDFTWIKNPPRAIRLTLPQLEIRGSYNLQ D	LLAQAKLSTL
13	335	1.01e-47	PLSANGYL LL	IQPHHTLDRKVEALIFQINFLTRMKNLSPRAIRLTVPQLTLKASYDL QD	LLAQAKLPTL
9	335	1.72e-46	PLGERATL LL	IQPHCTSDLRVEALIFRNDLLTWIENPPRAIRLTLPQLEIRGSYNLQ D	LLAEDKLPTL

Figure 1.3: Block Diagram of Antithrombotic peptide

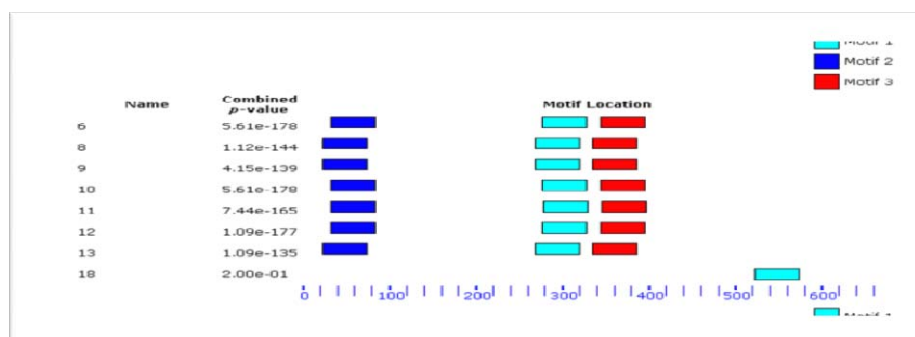


Table1.2:-Sites of Motif one in different antithrombotic peptides.

Name	Start	p-value	Sites		
10	277	8.76e-66	TGASVDSTLA	FNTYVHFQGKMKGFSLLAEFQEFWVDNSTSVSVFMLS G M G T F Q H W S D I Q D	NFSVTQVPFT
6	277	8.76e-66	MGASVDSTLA	FNTYVHFQGKMKGFSLLAEFQEFWVDNSTSVSVFMLS G M G T F Q H W S D I Q D	NFSVTQVPFT
11	278	2.76e-65	TGASADSNLV	FNTYVHFQGKMKGFSLLAEFQEFWVDNSTSVSVFMLS G M G T F Q H W S D A Q D	KFSVTQVPFT
12	277	1.77e-64	MGASVDSTLA	FNTYVHFQGKMKGFSLLAEFQEFWVDNSTSVSVFMLS G M G T F Q H W S D V Q D	NFSVTQVPFT
8	268	4.71e-58	EGVSTDSTLF	FNTYVHFQGKMKGFSLLAEFQEFWVDNSTSVSVFMLS G T G N F Q H W S D A Q N	NFSVTRVPLG
9	268	6.42e-55	EGVSTDSTLL	FNTYVHFQGTMRGFSLLAEFQEFWVDNSTSVSVFMLS G T G N F Q H W S D A Q N	NFSVTCVPLG

Table1.3: motif information with sequence logo and regular expression present in antithrombotic peptides

Sr. No.	Motif No	E-Value	Width	Entropy	Sites	Sequence logo	Regular expression
1	1	.9e-219	50	199.3	7		FNTYVHFQGKM[K R]GFS[LQ]L[AP][EG][PL][QH]EFWVDNSTSVSPMLSG [MT]G[TN]FQHWS D [AI]Q[DN]
2	2	2.1e-181	50	186.3	7		GDRVYIHPFHL[V L][IY]H[NS][EK]STC[EA]QL[AE][KN][AP][NS][AV][GE][KT][P L][KP][DE]PTF [IET]P[AV]PIQAK[T S]S P VDEK
3	3	1.8e-173	50	183.5	7		IQPH[YC][AT]SDLD [KR]VE[GA]L[TI]FQ [QH][ND][SF]L[NT] W[M]K[KN][LP][SP] PR[AT][I HR]LT [ML] QL [VE][LI][QR] GS Y [DN]L QD

Opioid peptides

type of sequence	protein
number of sequences	34
shortest sequence (residues)	5
longest sequence (residues)	268
average sequence length (residues)	154.8
total dataset size (residues)	5264

Figure 1.4: MEME output. Block diagram of motif present in Opioid peptides. This portion of an MEME HTML output form shows a protein motif that MEME has discovered in the input sequences.

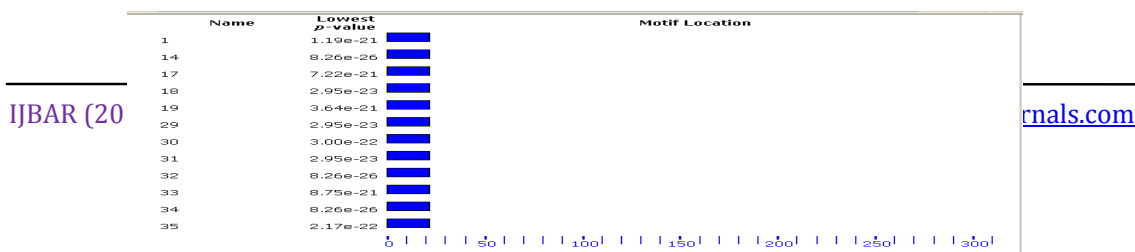


Table1.4: The sites identified as belonging to the motif are indicated, and above them is the ‘consensus’ of the motif and a color-coded bar graph showing the conservation of each position in the motif present in Opioid peptides.

Name	Start	p-value	Sites
34	0	8.26e-26	MKLLILTCLVAAVALARPKHPI KHQGLPQEV
32	0	8.26e-26	MKLLILTCLVAAVALARPKHPI KHQGLSSEVL
14	0	8.26e-26	MKLLILTCLVAAVALARPKHPI KHQGLPQEV
31	0	2.95e-23	MKFFIFTCLLAVALAKNTMEH VSSSEESIIS
29	0	2.95e-23	MKFFIFTCLLAVALAKNTMEH VSSSEESIIS
18	0	2.95e-23	MKFFIFTCLLAVALAKNTMEH VSSSEESIIS
35	0	2.17e-22	MKLLILTCLVATALARHKFHL GHLKLTQEQP
30	0	3.00e-22	MRLLILTCLVAAVALARPKLPL RYPERLQNPS
1	0	1.19e-21	MKLFIFTCLLAVALAKHKSEQ QSSSEESVSI
19	0	3.64e-21	MKLLILTCLVASAVAMPKFPF RHTELFQTQR
17	0	7.22e-21	MKLLILTCLVAAALALPRAHR RNAVSSQTQQ
33	0	8.75e-21	MKLLILTCLVAAAFAMRLHS RNAVSSQTQQ

Figure1.5: combined block diagram of Opioid peptides

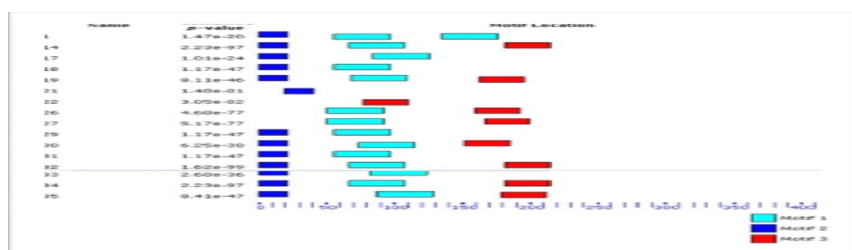


Table1.5: motif information of opioid peptides with sequence logo and regular expression

Sr. No.	Motif No	E-Value	Width	Entropy	Sites	Sequence logo	Regular expression
1	1	7.3e-130	43	114.9 (bits)	13		[QE][AV][MV][ER][DN]A[KN][QE][ME][EK][AY][SG][SI][SG]SSSEE[IS][VA][PE][NV][SA][TA]E[QE][KV][YHK][QT][KV][ED]D[VK][PH][SY][QE][RK][YA]L
2	2	1.7e-116	21	186.3 (bits)	12		MK[LF][LF][LF]TC L[VL]AVALA[RK][P N][KT][HM][PEH][H I]
3	3	9.0e-090	32	183.5 (bits)	6		PLGTQYTDAPSFSD IPNPIGSENS[EG]K[IT]TMPLW




Motif one present *in* opioid peptides (32, 27, 26, 34, 14, 31, 29, 18, 19, 35, 33)

Motif two present *in* opioid peptides (34, 32, 14, 31, 29, 18, 35, 30, 1, 19, 17, and 33)

Motif three present *in* opioid peptides 32, 27, 26, 34, 14, 30

Mineral binding peptides	
type of sequence	protein
number of sequences	10
shortest sequence (residues)	105
longest sequence (residues)	1464
average sequence length (residues)	597.2
total dataset size (residues)	5972

Table 1.6: motif information of opioid peptides with sequence logo and regular expression

<i>Sr . No.</i>	<i>Moti f No</i>	<i>E- Value</i>	<i>Wi dth</i>	<i>Sit es</i>	<i>Sequence logo</i>	<i>Regular expression</i>
1	1	2.4e+000	21	3		[TC][INP][YD][CD]F [CSW][SN][NT]E[KNV][HN Q][TC[CST][VC][IKR][AQR][KLQ][EWY][AI][DEQ]F
2	2	9.8e+000	8	2		QY[HM][IM][AP]C[NP]Y
3	3	1.0e+001	11	2		[CV][CW]WV[AT][IL]TW[F H]

Name	Start	p-value	Sites
7	404	2.30e-23	LIHFQGGFFVA TIYCF CNNEVQTTVKRQWQF KIQWNQRWGR
2	281	3.41e-21	QSSRHRRALD TNYCF SSTEKNCCVRQLYIDF RKDLGWKWIH
5	584	1.30e-18	DETDASACNK CPDDF WSNEN TSCIA KEIEF LSWTEPFGIA

Figure1.6: site present in motif of mineral binding peptides

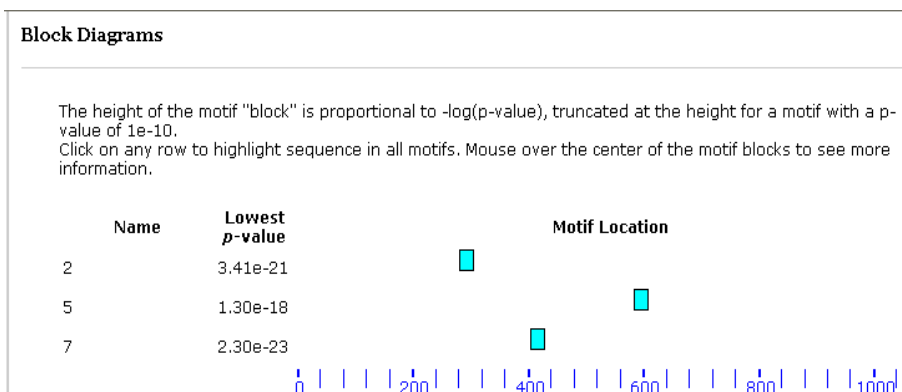


Figure1.7: Block diagram of motif of mineral binding peptides

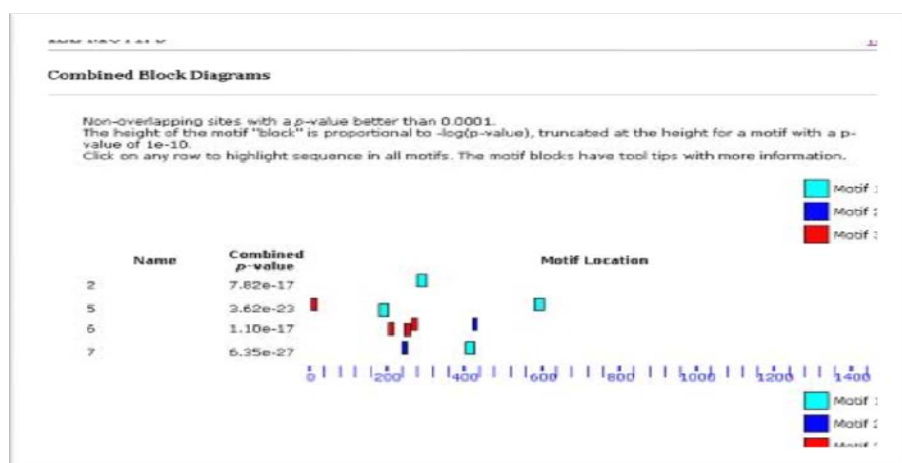


Figure1.7: Block diagram of motif of mineral binding peptides

Motif one present *in* mineral peptides 2, 5 and 7.

Motif two present *in* mineral peptides 6 and 7.

Motif three present *in* mineral peptides 5 and 6.

Immunomodulatory peptides

type of sequence	protein
number of sequences	11
shortest sequence (residues)	18
longest sequence (residues)	771
average sequence length (residues)	219.6
total dataset size (residues)	2416

Sr. No.	Motif No.	E-Value	Width	Sites	Sequence logo	Regular expression
1	2	$2.6e+001$	8	2		H[ES][FW]T[KW]PEN
2	1	$1.0e-005$	13	2		GCWTKSYPPKPCF
3	3	$6.0e+001$	32	2		FT[IV][KM]ET[EV]C[PT][PR][ST][AT][QR][LQ][PS][PV][DE]DCDFK[EK][DN]G[LV][IV][KR][DR]C[MS]G

Table1.6 : regular expression, sequence logo, sites along with e-value of motif one of Immunomodulatory peptides

Name	Start	p-value	Sites
1	78	2.72e-36	GDPDTPKPV FTVKETVCRRTTQQSPEDCDFKKDGLVKRCMG TVTLNQARG
5	67	5.40e-35	VDLSTLRALN FTIMETECTPSARLPVDDCDFKENGVIKDCSG PVSVLQDTPE

Figure1.7: sites present in motif one of immunomodulatory peptides

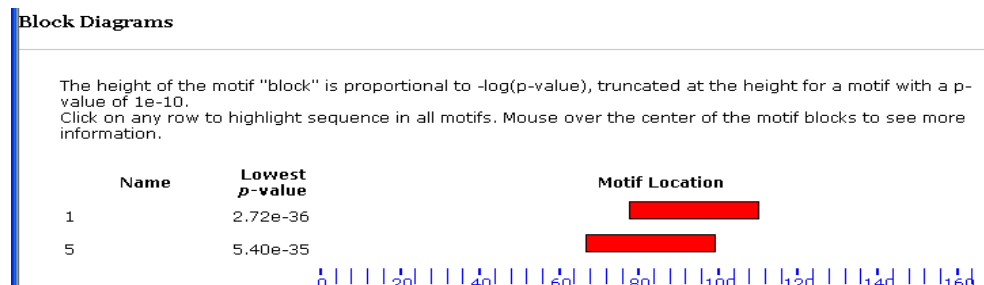


Figure1.7: Block diagram of motif one of immunomodulatory peptides

Motif one present in Immunomodulatory peptides are 3 and 4.

Motif two present in Immunomodulatory peptides 2 and 9.

Motif three present in Immunomodulatory peptides 1 and 5.

Combined Peptides

type of sequence	protein
number of sequences	111
shortest sequence (residues)	5
longest sequence (residues)	1464
average sequence length (residues)	223.5
total dataset size (residues)	24806

Sr. No.	Motif No	E-Value	Entropy	Width	Sites	Sequence logo	Regular expression
1	1	9.2e-218	199.4 (bits)	50	7		FNTYVHFQGKM[K R]GFS[EG][PL][QH] EFWVDNSG[MT]G[TN]FQHWSD[AI]Q[DN]
2	2	1.5e-175	191.4 (bits)	50	7		AGDRVYIHPFHL[V L][IY]H[NS][EK]ST C[EA]QL[AE][KN][A P][NS][AV]GE][KT][PL][KP][DE]PTF[IE T][AV]PIQAK[TS]SP VDE
3	3	1.5e-163	181.0 (bits)	50	7		LIQPH[YC][AT]SDL D[KR]VE[GA]L[TI]F Q[QH][ND][SF]L[NT]W[MI]K[KN][LP][S P]PR[AT][I][HR]LT[M L]PQL [VE][LI][QR]GSY[D N]LQ

Table1.7 : regular expression, sequence logo, sites along with e-value of motif one when all peptide analysed together

Name	Start	p-value	Sites	
12	343	1.68e-62	VPFTESACLL	LIQPHYASDLKVEGLTFQQNSLNWMKKLSRAIHLTMQLVLQGSYDLQDLLAQAEIPA
10	343	1.44e-61	VPFTESACLL	LIQPHYASDLKVEGLTFQQNSLNWMKKLSRTIHLTMQLVLQGSYDLQDLLAQAEIPA
6	343	1.44e-61	VPFTESACLL	LIQPHYASDLKVEGLTFQQNSLNWMKKLSRTIHLTMQLVLQGSYDLQDLLAQAEIPA
11	344	1.53e-58	VPFTESACLL	LIQPHYASDLKVEGLTFQQNSLNWMKKLSRAIHLTMRLVLRGSYDLQDLLAQAEIPT
8	334	7.91e-48	VPLGESVTLL	LIQQCASDLDRVEVLVFQHDFTWIKNPPRAIRLTLQLEIRGSYNLQDLLAQAKLST
13	334	4.33e-47	VPLSANGYLL	LIQPHETDLRKVEALIFQHNFLTRMKNLSPRAIHLTVQLTLKASYDLQDLLAQAKLPT
9	334	7.56e-46	VPLGERATLL	LIQPHCTSDLDRVEALIFRNDLLTWIENPPRAIRLTLQLEIRGSYNLQDLLAEDKLPT

Table 1.8 : sites present in motif one when all peptide analysed together

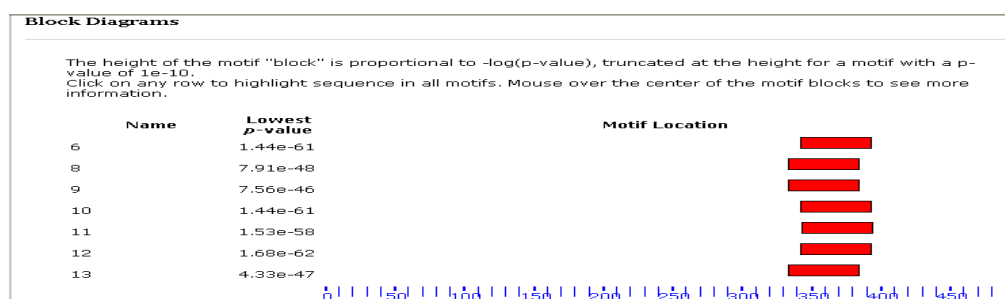


figure 1.7 : Block diagram of motif one when all peptide analysed together

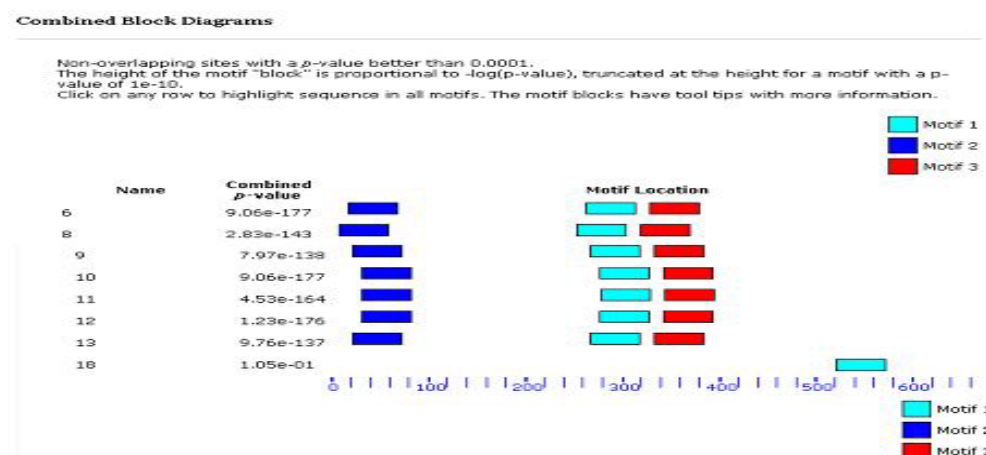


figure 1.8 : Block diagram of motif one when all peptide analysed together

Motif one present in combined peptides is 10, 6, 11, 12, 8, 9 and 15.

Motif two present in combined peptides 12, 10, 6, 11, 8, 9 and 13.

Motif three present in combined peptides 12, 10, 6, 11, 8, 9 and 13.

Casein Peptides

type of sequence	protein
number of sequences	37
shortest sequence (residues)	12
longest sequence (residues)	415
average sequence length (residues)	150.0
total dataset size (residues)	5549




Sr. No.	Motif No	E-Value	Width	Sites	Sequence logo	Regular expression
1	1	7.3e-130	41	11		[QE][AV][MV][ER][DN]A[KN][QE][ME][EK][AY][SG][SI][SSSEE][IS][VA][PE][NV][SA][TA]E[QE][KV][YKH][QT][KV][ED][VK][PH][SY][QE][RK][YA]L
2	2	1.7e-116	21	12		MK[LF][LF]TCL[VL]AVALA[RK]PN][KT][HM][PEH][HI]
3	3	9.0e-090	32	6		PLGTQYTDAPSFS DIPNPIGSENS[EG]K[IT]TMPLW

Table1.8 : regular expression, sequence logo, sites along with e-value of motif one of casein derived peptides

Name	Start	p-value	Sites
32	182	1.35e-41	AYPSGAWYYL FLGTQYTDAPSFSDIPNPIGSENSGKITMPLW
27	167	1.35e-41	AYPSGAWYYL FLGTQYTDAPSFSDIPNPIGSENSGKITMPLW
26	159	1.35e-41	AYPSGAWYYL FLGTQYTDAPSFSDIPNPIGSENSGKITMPLW
34	182	3.86e-41	AYPSGAWYYV FLGTQYTDAPSFSDIPNPIGSENSEKTTMPLW
14	182	3.86e-41	AYPSGAWYYV FLGTQYTDAPSFSDIPNPIGSENSEKTTMPLW
30	152	2.24e-21	AAYPYAVWYY QIMQYVPFPFSDISNPTAHENYEKNNVMLQ W

Table1.7 : sites of motif one of casein derived peptides

Block Diagrams

The height of the motif "block" is proportional to $-\log(p\text{-value})$, truncated at the height for a motif with a p-value of $1e-10$. Click on any row to highlight sequence in all motifs. Mouse over the center of the motif blocks to see more information.

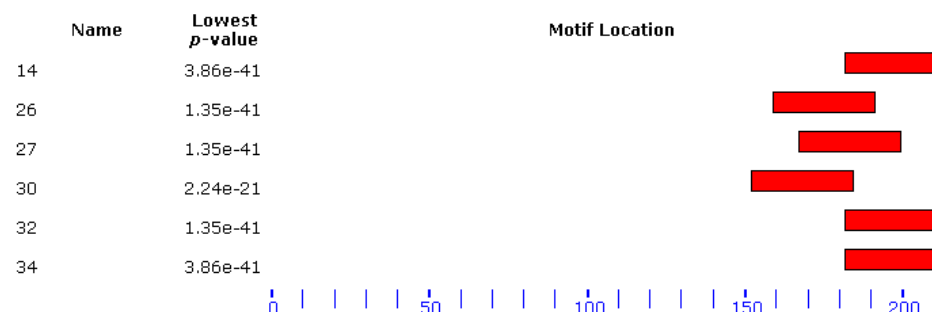


figure 1.9 : Block diagram of motif one casein derived peptides

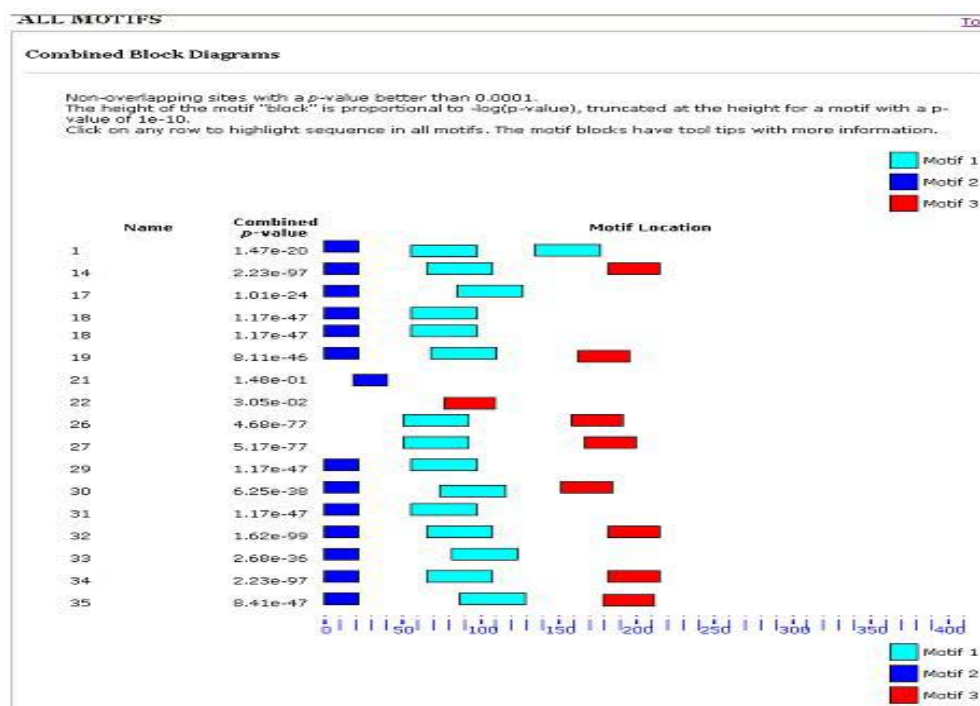


figure 1.10 : Block diagram of motifs present in casein derived peptides

Motif one present in Casein peptides are 32, 27, 26, 34, 14, 31, 29, 18, 19, 35 and 33

Motif two present in Casein peptides 1, 14, 17, 18, 19, 29, 30, 31, 32, 33, 34 and 35.

Motif three present in Casein peptides 32, 26, 27, 34, 14 and 30.

Conclusion

In genetics, a sequence motif is a nucleotide or amino-acid sequence pattern that is widespread and has, or is conjectured to have, a biological significance. For proteins, a sequence motif is distinguished from a structural motif, a motif formed by the three dimensional arrangement of amino acids, which may not be adjacent. The notation [XYZ] means X or Y or Z, but does not indicate the likelihood of any particular match. For this reason, two or more patterns are often associated with a single motif: the defining pattern, and various typical patterns. For example, the defining sequence for the IQ motif may be taken to be: [FILV] Qxxx[RK]Gxxx[RK]xx[FILVWY] where x signifies any amino acid, and the square brackets indicate an alternative (see below for further details about notation). Usually, however, the first letter is I, and both [RK] choices resolve to R. Since the last choice is so wide, the pattern IQxxxRGxxxR is sometimes equated with the IQ motif itself, but a more accurate description would be a consensus sequence for the IQ motif. De novo

computational discovery of motifs are very common today and there are software programs which, given multiple input sequences, attempt to identify one or more candidate motifs and MEME suite is most suitable as it generates statistical information for each sequence. As stated above MEME describes that although bioactive peptides precursors present in different source of origin, they contains common patterns of amino acids. It is again noted that motifs may overlap with one another due to the reason of common consensus patterns.

Acknowledgement: We are thankful to Mr. Gyanendar Singh, Research Fellow, Bioinformatics Facility, SVP University, Meerut for his valuable suggestions and help.

References

1. Kitts D. D., Weiler, K. A. Bioactive proteins and peptides from food sources Applications of bioprocesses used in isolation and recovery. Current

- Pharmaceutical Design, 2003, 9, pp. 1309–1323.
2. Bailey, T. L. and Elkan, C. Unsupervised Learning of Multiple Motifs In Biopolymers Using EM. *Mach. Learn.*, 1995, 21, 51–80.
 3. Bailey, T. L. and Elkan, C. The value of prior knowledge in discovering motifs with MEME. In Rawlings, C., Clark, D., Altman, R., Hunter, L., Lengauer, T. and Wodak, S. (eds), In *Proceedings of the Third International Conference on Intelligent Systems for Molecular biology*, July. AAAI Press, Menlo Park, CA, 1995, pp. 21–29.
 4. FitzGerald R., Murray, B. A. Bioactive peptides and lactic fermentations. *International Journal of Dairy Technology*. 2006, 59, pp. 118–125.
 5. Korhonen, H. J. T., Pihlanto-Leppala, A. Bioactive peptides: production and functionality. *International dairy journal*, 2006, 16, 945–960.
 6. Van Beresteijn E. C. H., Peeters, R. A., Kaper, J., Meijer, R. J. G., Robbeb, A. J. P. M. and Schmidt, D. G. Molecular mass distribution, immunological properties and nutritive value of whey protein hydrolysates. *Journal of Food Science*, 1994, 57, 619 – 625.
 7. FitzGerald R.J., Murray, B.A. and Walsh D. J. Hypotensive peptides from milk proteins. *Journal of Nutrition*, 2004, 134, 980S–988S.
 8. Jolles, P., Levy-Toledano, S, Fiat, A. M., Soria, C., Gillessen, D, Thomaidis, A. “Analogy between fibrinogen and casein Effect of an undecapeptide isolated from kappa-casein on platelet function”. *European Journal of Biochemistry*, 1986, 158, 379–382.
 9. Jollès, P. and Caen, J.P. Parallels between milk clotting and blood clotting: opportunities for milk-derived products. *Trends in Food Science and Technology*, 1991, 2, 42–43.
 10. Jollès, P. and Henschen, A. Comparison between the clotting of blood and milk. *Trends in Biochemical Science*, 1982, 7, 325–328.
 11. Chabance, B., Jollès, P., Izquierdo, C., Mazoyer, E., Francoual, C., Drouet, L. and Fiat, A. M. Characterization of an antithrombotic peptide from κ -casein in newborn plasma after milk ingestion". *British Journal of Nutrition*, 1995, 73, 583–590.