Available online at **www.ijntps.org** | **ISSN:** 2277 – 2782
INTERNATIONAL JOURNAL OF NOVEL TRENDS IN PHARMACEUTICAL SCIENCES

RESEARCH ARTICLE

Identification of single de novo drug candidate for dengue and filaria on *Aedes aegypti* and *Culex quinquefasciatus* mosquitoes using insilico Protocols

Cardilya Grace. C[1*], *Balaji. M*[2] *and John William. S*[1]

[1]Dept. of Advanced Zoology and Biotechnology, Loyola College, Chennai-34.
[2]Director, Akshaya Neuroinformatics, Research Center, Chennai-5.

Article Info

Article history
Received 12 May 2012
Revised 16May 2012
Accepted 10 Jun 2012
Available online 30 Jun 2012

Keywords
Protein Sequence Analysis, Sequence Comparison and Motif Prediction.

Abstract

Aedes aegypti is a mosquito that can spread the dengue fever, Chikungunya and yellow fever viruses, and other diseases. The mosquito can be recognized by white markings on legs and a marking in the form of a lyre on the thorax. The mosquito originated in Africa but is now found in tropical and subtropical regions throughout the world. *Culex quinquefasciatus* is the vector of lymphatic filariasis caused by the nematode *Wuchereria bancrofti* in the tropics and sub tropics. The present research investigation studies primarily focus on the molecular similarities of *Aedes aegypti* and *Culex quinquefasciatus*. Our aim is to reduce the pathological effects of dengue and filarial disease in humans. The molecular sequence alignment is one of the best drug targets for structure based drug designing. The protein functional similarities identification using advanced motif prediction tool MEME. We finally conclude from the results that the identified conserved motif sequences are the potential protein target sequences for structure based drug designing and molecular drug docking studies. This work would definitely be useful in the field of Clinical Pathology, Computational Entomology and Cheminformatics.

INTRODUCTION

Aedes aegypti is a mosquito that can spread the dengue fever, Chikungunya and yellow fever viruses, and other diseases. The mosquito originated in Africa but is now found in tropical and subtropical regions throughout the world. It belongs to the tribe Aedini of the dipteran family Culicidae and to the genus *Aedes* and subgenus *Stegomyia*. It infects 50 to 100 million people worldwide a year, leading to half a million hospitalizations, and approximately 12,500–25,000 deaths. Symptoms include fever, headache, muscle and joint pains, and a characteristic skin rash that is similar to measles. In a small proportion of cases the disease develops into the life-threatening dengue hemorrhagic fever, resulting in bleeding, low levels of blood platelets and blood plasma leakage, or into dengue shock syndrome.[1] Apart from eliminating the mosquitoes, work is ongoing on a vaccine, as well as medication targeted directly at the virus.

Culex quinquefasciatus is the vector of lymphatic filariasis caused by the nematode *Wuchereria bancrofti* in the tropics and sub tropics. Filariasis is "considered" endemic in tropical and subtropical regions of Asia, Africa, Central and South America, and Pacific Island nations, with more than 120 million people infected and one billion people at risk for infection. The most spectacular symptom of lymphatic filariasis is elephantiasis—edema with thickening of the skin and underlying tissues—which was the first disease discovered to be transmitted by mosquito bites. The strategy for eliminating transmission of lymphatic filariasis is mass distribution of medicines that kill the microfilariae and stop transmission of the parasite by mosquitoes in endemic communities [2].

RESEARCH METHODS

Two software tools were used to identify the motif and conserved regions present in the target sequences. The motif studies were performed using **MEME** tool in order to identify the molecular conserved motif present in the Aedes aegypti and Culex quinquefasciatus. The sequence alignment studies were done using **T COFFEE** server in order to predict the conserved amino acids present in the species (*Aedes aegypti* and *Culex quinquefasciatus*). The results obtained are as follows.

▼ To whom correspondence should be addressed:
Ms. Cardilya Grace. C
E-Mail address: cardilya@gmail.com

>gi|157109037|ref|XP_001650496.1| cdk10/11 [Aedes aegypti]
MTKDKRKAEHDGKQVTDKVAVGSGSSDPVGPIVRKGALMSFRSKKFTNIPIKDMYGRCRYVNAFMKCNRV
GEGTYGIVFRARDTENEEIVALKKVRIDQEMFKDGFPVSGLREIQILKNCNHENVVKLKEVVVGNSLESI
FLVMEFCEQDLASLLDNMETPFSESQVKCIVNQLLKGLKYLHSQFIIHRDLKVSNLLLTDKGCLKIADFG
LARYISDSDKPMTPGLVTLWYRPPELLFGSKVQTTAVDMWATGCILGELLAHKPLLPGVSEISQIELIIE
LLGTPSETIWPDFSSLPAVQNFTLRSQPYNNLKPKFAWLSSAGLRLLNFLFMYDPKKRATAEECLQSSYF
KEAPLPCDPKLMPTFPHHRELKNTAKEPTEISASTDQITIPTISDLLGSLVKKRRTE

>gi|170030910|ref|XP_001843330.1| cell division protein kinase 10 [Culex quinquefasciatus]
MTKEKRKAADGGKQTGTSDPCGPITRKGLMSFLTKQFTDIPSKDIFGKCRYVSAFQKCNRVGEGTYGIVF
RARDTTNNEIVALKKVRIDQEIFKDGFPVSGLREIQILKSCNHENVVQLKEVVVGNSLESIFLVMEFCEQ
DLASLLDNMESPFTESQVKCIVIQLLKGLRYLHANFIIHRDLKVSNLLLTDKGCLKIADFGLARYQSDST
KPMTPGLVTLWYRSPELLFGAKEQTTAVDMWAAGCILGELLAHKPLLPGVSEISQIELIIDLLGTPSETI
WPDFSRLPALQNFTLKAQPYNNLKPKFAWLSSAGLRLLNFLFMYDPKKRASAEECLQSSYFKEAPLPCDP
KLMPTFPHHRDLKNAPSAEAPSSSSNVNAIFDQTTMPTISDLLGSLVKKRRID

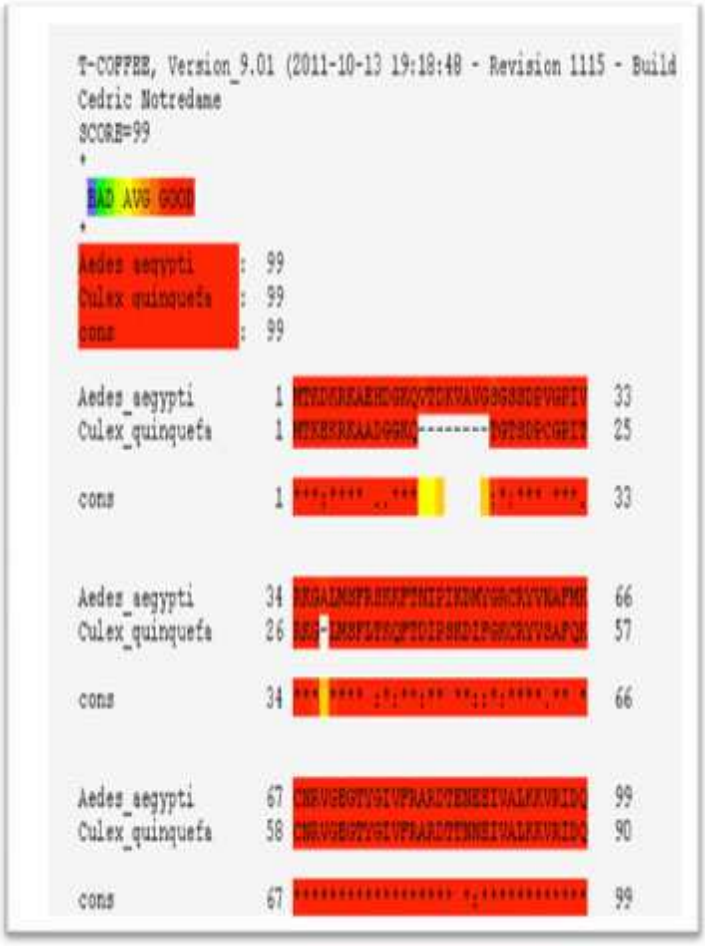
Figure : 1 Sequence Similarity Searches - BLAST Result

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
XP_001843330.1	cell division protein kinase 10 [Culex quinquefasciatus] >gb|EDS3219	834	834	100%	0.0	100%
XP_001650496.1	cdk10/11 [Aedes aegypti] >gb|EAT43373.1| cdk10/11 [Aedes aegyp	704	704	100%	0.0	83%
XP_315879.4	AGAP005851-PA [Anopheles gambiae str. PEST] >gb|EAA11953.5| AG	610	610	95%	0.0	75%
XP_003489943.1	PREDICTED: cyclin-dependent kinase 10-like [Bombus impatiens]	604	604	99%	0.0	72%
XP_003392867.1	PREDICTED: cyclin-dependent kinase 10-like [Bombus terrestris]	603	603	99%	0.0	72%
XP_392973.4	PREDICTED: cyclin-dependent kinase 10 [Apis mellifera]	600	600	99%	0.0	72%
NP_001154939.1	cyclin-dependent kinase 10 isoform 2 [Nasonia vitripennis]	598	598	100%	0.0	71%
NP_001154938.1	cyclin-dependent kinase 10 isoform 1 [Nasonia vitripennis]	597	597	99%	0.0	72%
XP_974492.1	PREDICTED: similar to cdc2-related kinase [Tribolium castaneum] >gb	585	585	95%	0.0	71%
EGI69589.1	Cell division protein kinase 10 [Acromyrmex echinatior]	583	583	96%	0.0	71%
EFN66919.1	Cell division protein kinase 10 [Camponotus floridanus]	582	582	96%	0.0	71%
EHJ70824.1	cdc2-related kinase [Danaus plexippus]	580	580	95%	0.0	71%
ADD20306.1	cell division protein kinase 10 [Glossina morsitans morsitans]	575	575	96%	0.0	70%
NP_001037345.1	cdc2-related kinase [Bombyx mori] >dbj|BAA21484.1| cdc2-related k	575	575	95%	0.0	71%
EFR30000.1	hypothetical protein AND_00677 [Anopheles darlingi]	575	575	98%	0.0	71%
XP_002428425.1	mitogen-activated protein kinase ERK-A, putative [Pediculus humanus	568	568	95%	0.0	69%
AAC79672.3	putative cdc2-related kinase [Haematobia irritans irritans]	569	569	95%	0.0	69%

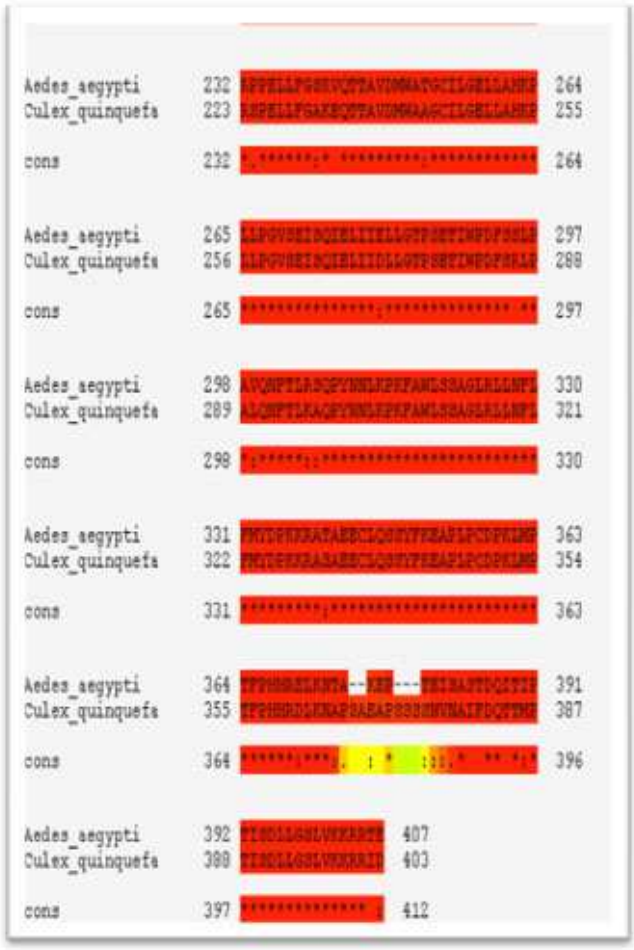
Aedes aegypti* and *Culex quinquefasciatus

Figure: 2 Sequence Alignment – TCOFFEE Server



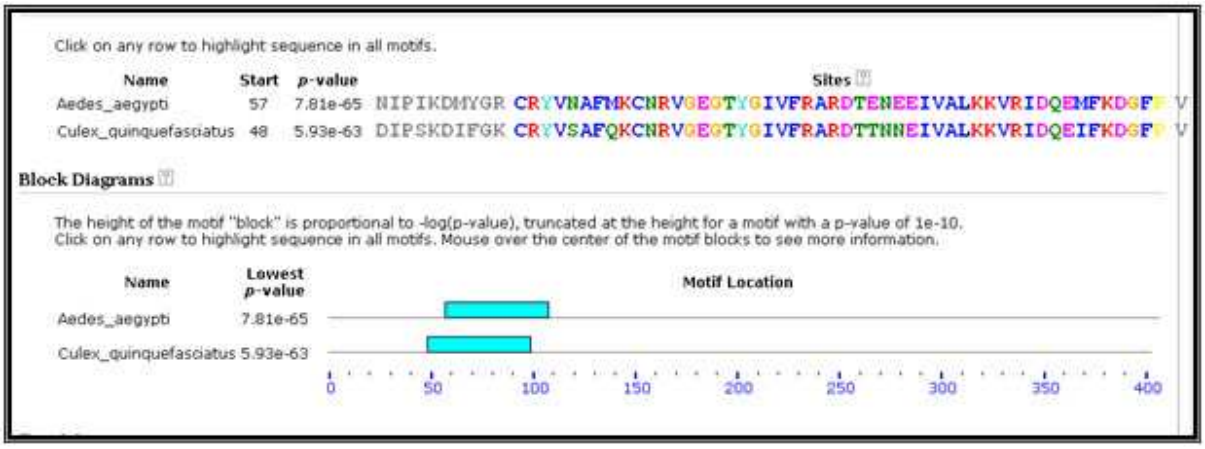
Red Color – Conserved regions (99%) - *Aedes aegypti* and *Culex quinquefasciatus*

Figure: 2.2 Sequence Alignment – TCOFFEE Server



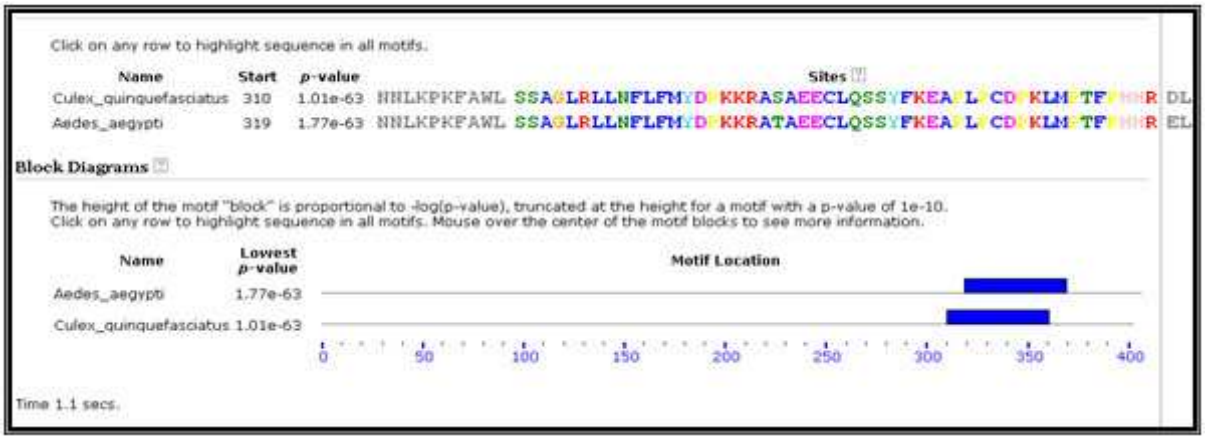
Red Color – Conserved regions (99%) - *Aedes aegypti and Culex quinquefasciatus*

Figure: 3 Motif Prediction –MEME Server



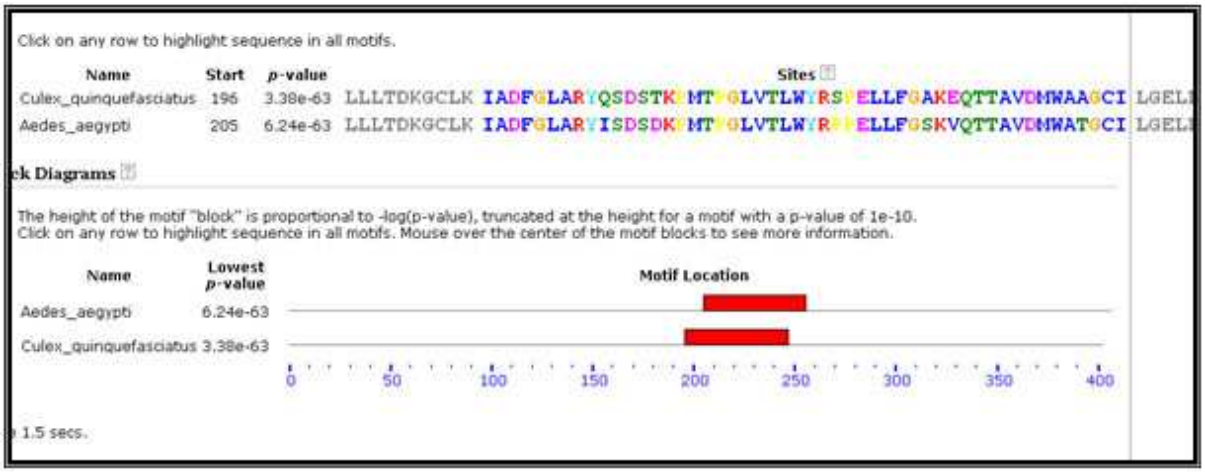
Light Blue color indicates molecular conserved Motif - *Aedes aegypti* and *Culex quinquefasciatus*

Figure: 3.1 Motif Prediction –MEME Server



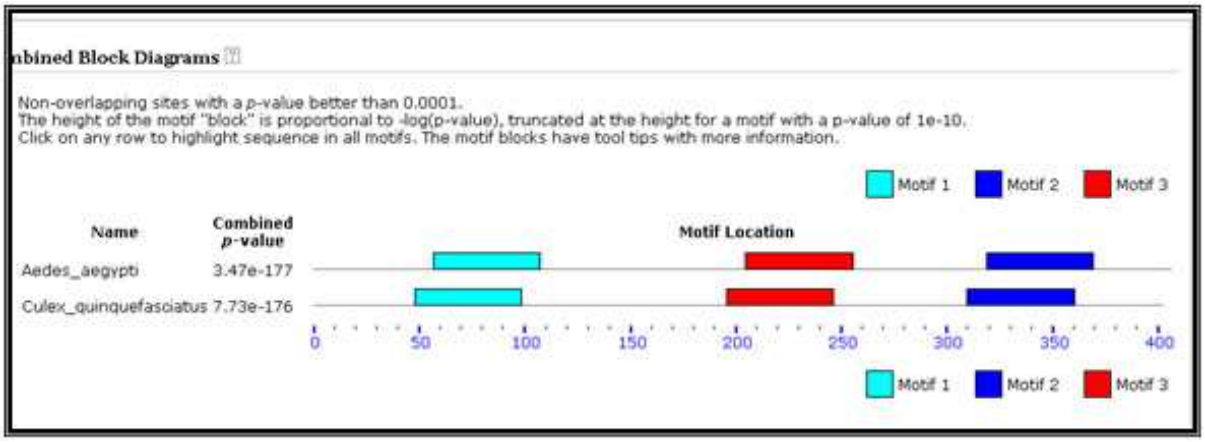
Dark Blue color indicates molecular conserved Motif - *Aedes aegypti* and *Culex quinquefasciatus*

Figure: 3.2 Motif Prediction –MEME Server



Red color indicates molecular conserved Motif - *Aedes aegypti* and *Culex quinquefasciatus*

Figure: 3.3 Motif Prediction –MEME Server



Light Blue, Red and Dark Blue color indicates molecular conserved Motif - *Aedes aegypti* and *Culex quinquefasciatus*

RESULT AND CONCLUSION

The sequence alignment was done using TCOFFEE server [3]. This software describes the Pair wise sequence alignment methods which are used to find the best-matching piecewise (local) or global alignments of two query sequences. Pairwise alignments can only be used between two sequences at a time, but they are efficient to calculate and are often used for methods that do not require extreme precision (such as searching a database for sequences with high similarity to a query).

The three primary methods of producing pairwise alignments are dot-matrix methods, dynamic programming, and word methods [4] however; multiple sequence alignment techniques can also align pairs of sequences. Although each method has its individual strengths and weaknesses, all three pairwise methods have difficulty with highly repetitive sequences of low information content - especially where the number of repetitions differ in the two sequences to be aligned. One way of quantifying the utility of a given pairwise alignment is the 'maximum unique match' (MUM), or the longest subsequence that occurs in both query sequence. Longer MUM sequences typically reflect closer relatedness.

We are going to describe the T-coffee tool which is used to analyze the sequences of *Aedes aegypti* and *Culex quinquefasciatus*. **(Fig.2-2.2).** (T-Coffee) for multiple sequence alignment that provides a dramatic improvement in accuracy with a modest sacrifice in speed as compared to the most commonly used alternatives. The method is broadly based on the popular progressive approach to multiple alignments but avoids the most serious pitfalls caused by the greedy nature of this algorithm. With T-Coffee we pre-process a data set of all pair-wise alignments between the sequences. This provides us with a library of alignment information that can be used to guide the progressive alignment. Intermediate alignments are then based not only on the sequences to be aligned next but also on how all of the sequences align with each other.

This alignment information can be derived from heterogeneous sources such as a mixture of alignment programs and/or structure superposition [6]. Here, we illustrate the power of the approach by using a combination of local and global pair-wise alignments to generate the library. The resulting alignments are significantly more reliable, as determined by comparison with a set of 141 test cases, than any of the popular alternatives that we tried. The improvement, especially clear with the more difficult test cases, is always visible, regardless of the phylogenetic spread of the sequences in the tests.

The generation of a protein sequence is much easier than the determination of a protein structure. However, the structure of a protein gives much more insight in the function of the protein than its sequence. Therefore, a number of methods for the computational prediction of protein structure from its sequence have been developed**.** *Ab initio* prediction methods use just the sequence of the protein. Threading and Homology Modeling methods can build a 3D model for a protein of unknown structure from experimental structures of evolutionary related proteins.

MEME

De novo computational discovery of motifs

We used motif prediction for species *Aedes aegypti* and *Culex quinquefasciatus* by using the MEME tool [6,7,9,10,11]. There are software programs which, given multiple input sequences, attempt to identify one or more candidate motifs. One example is MEME, which generates statistical information for each candidate. Other algorithms include AlignAce, Amadeus, CisModule, FIRE, Gibbs Motif Sampler, PhyloGibbs, and Weeder. SCOPE is an ensemble motif finder that uses several algorithms simultaneously. PMS or PMS is another motif discovery program that is based on combinational approach. There currently exist more than 100 publications with similar algorithms without a comprehensive benchmark so selecting one is not straightforward. **Fig (3-3.3)** shows the various motif sequences present in the selected species (*Aedes aegypti* and *Culex quinquefasciatus*). All the above results were discussed.

In bioinformatics, a sequence alignment is a way of arranging the sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences. Aligned sequences of nucleotide or amino acid residues are typically represented as rows within a matrix. Gaps are inserted between the residues so that identical or similar characters are aligned in successive columns.

The potential conserved sequence regions were identified. Both the sequence and the function are conserved. These are the evidences for designing the same de novo drug candidate for curing the Dengue and Filaria. We finally conclude from the results that the identified conserved motif sequences are the potential protein target sequences for structure based drug designing and molecular drug docking studies. This work would definitely

be useful in the field of Clinical Pathology, Computational Entomology and Cheminformatics.

Reference

1. Ranjit S. and Kissoon N. Dengue hemorrhagic fever and shock syndrome. *Pediatr. Crit. Care Med.* 2010, 12(1), 90–100. doi:10.1097/PCC.0b013e3181e911a7.
2. The Carter Center. *Summary of the Third Meeting of the International Task Force for Disease Eradication*, 2002, retrieved 2008-07-17.
3. Notredame C, Higgins D.G., Heringa J.T.-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol, 2000, 302(1): 205-17.
4. Mount D.M. *Bioinformatics: Sequence and Genome Analysis*. Cold Spring Harbor Laboratory Press. 2004. 2nd edition.
5. Zhang Y. Progress and challenges in protein structure prediction. *Curr Opin Struct Biol,* 2008, 18(3), 342–348.DOI:10.1016/j.sbi.2008.02.004.
6. Timothy L.B. and Charles E. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, 1994, 28-36.
7. William N.G., Timothy L.B., Charles P.E. and Michael E.B. Hidden Markov Model Analysis of Motifs in Steroid Dehydrogenases and their Homologs. Biochemical and Biophysical Research Communications, 1997, 231, 760-766,
8. William N.G., Timothy L.B., Charles P.E. and Michael E.B. Meta-MEME: Motif-based Hidden Markov Models of Protein Families. *Computer Applications in the Biological Sciences* . 1997, 13(4), 397-406.
9. Michael E.B., William N.G. and Charles P.E. Spinach CSP41, an mRNA-binding protein and ribonuclease, is homologous to nucleotide-sugar epimerases and hydroxysteroid dehydrogenases. Biochemical and Biophysical Research Communications, 1998, 248(2), 250-254.
10. Michael E.B., William N.G., and Charles P.E. A common ancestor for a subunit in the mitochondrial proton-translocating NADH: ubiquinone oxidoreductase (complex I) and short-chain dehydrogenases/reductases. Cellular and Molecular Life Sciences, 1999, 55(3), 450-455,
11. John H, Charles G, William S.N., and Timothy L. B. Assessing phylogenetic motif models for predicting transcription factor binding sites. Bioinformatics (Proceedings of the Intelligent Systems for Molecular Biology Conference), 2009, 25(12), 339—347.