Journal of Advanced Bioinformatics Applications and Research.
Vol 1, Issue 1, June 2010, pp 11-16 http://www.bipublication.com

IDENTIFICATION OF DRUGGABLE (TARGET) PROTEIN OF HUMAN COUNTERPART IN *C. elegans*

Mukta C. Zarkar[1], Swati V. Ghuge[1], Alok Shekhar[#]

[1]MGM'S College of CS & IT, Department of Biotechnology & Bioinformatics Nanded (MS) – 431605
Yeshwant College of IT, bioinformatics & Biotechnology, Parbhani (MS) – 431401

ABSTRACT

To design the drug from any protein it's very important to know about the Active site of protein or signaling pathway of that protein. If we know the active site of that protein then it is easy to design the drug, which can easily bind to that active site, or if we know the signaling pathway, then we can block the signaling pathway. If we block the signaling pathway then signals can't pass into the cell and disease will not caused. So, it is important to find the druggable proteins (cysteine protease, phosphates and kinase) and to study their signaling process. But firstly we have to know the function of these proteins, knowing the exact functions of these signaling proteins in humans is quite difficult, so we need another maodel organism for study. E.g. *C.elegans* is a small invertebrate organism having similar druggable proteins (protease, phosphatase and kinase) sequence to human's counterpart. As it is small organism, we can easily study different types of proteins interaction more easily We find the druggable protein sequences by using different tools like clustal X, tree view and HMMER. ClustalX is a widely used program for multiple sequence alignment. It performs multiple sequence alignment and gives the output in .aln and .dnd format.

Keywords: - Active site, Signaling pathway, Druggable, HMMER, clustalX.

Introduction:-

The functionality of any protein is maintained by the cell-signaling pathway. On the surface of cell membrane some proteins are present due to which the diseases are caused, by studying these proteins we can design the drug that can block the signaling pathway of these proteins, so these are called as druggable proteins. The examples of these druggable proteins are kinase, protease and phosphates. Knowing the exact functions of these signaling proteins in humans is quite difficult, so we need another model organism for study ex. C. elegans, by which we can study the nature and functions of these proteins.

Caenorhabditis elegans is a free-living, transparent nematode (roundworm), about 1 mm in length,[1] which lives in temperate soil environments. Research into the molecular and developmental biology of *C. elegans* was begun in 1974 by Sydney Brenner and it has since been used extensively as a model organism.[2]

C.elegans is a small invertebrate organism having similar druggable proteins (protease, phosphatases and kinase) sequence to human's counterpart. In this project, we can found these druggable protein sequences by using different tools like clustalX, Tree View and HMMER.

As *C.elegans* is very small organism so we can easily study the signaling

proteins pathway and we can also study different types of protein interaction more easily.

Biology

C. elegans is unsegmented, vermiform, and bilaterally symmetrical, with a cuticle integument, four main epidermal cords and a fluid-filled pseudocoelomate cavity. Members of the species have many of the same organ systems as other animals. In the wild, they feed on bacteria that develop on decaying vegetable matter. *C. elegans* has two sexes: hermaphrodites and males.[3] The sperm are stored in the same area of the gonad as the oocytes until the first oocyte pushes the sperm into the spermatheca (a kind of chamber where the oocytes become fertilized by the sperm).[4]

Dauer larvae can be transported by invertebrates including millipedes, insects, isopods, and gastropods. When they reach a desirable location they then get off, and at least in the lab they will also feed on the dead host if it dies.[5]

C. elegans is one of the few forms of life not known to have a natural virus.[6] Nematodes are capable of surviving desiccation, and in *C. elegans* the mechanism for this capability has been demonstrated to be Late Embryogenesis Abundant (LEA) proteins.[7]

Laboratory uses

From a research perspective, *C. elegans* has the advantage of being a multicellular eukaryotic organism that is simple enough to be studied in great detail. *C. elegans* is transparent, facilitating the study of cellular differentiation and other developmental processes in the intact organism. The developmental fate of every single somatic cell (959 in the adult hermaphrodite; 1031 in the adult male) has been mapped out [8][9]. These patterns of cell lineage are largely invariant between individuals, in contrast to mammals where cell development from the embryo is more largely dependent on cellular cues. In both sexes, a large number of additional cells (131 in the hermaphrodite, most of which would otherwise become neurons), are eliminated by programmed cell death (apoptosis).

Genome

C. elegans was the first multicellular organism to have its genome completely sequenced.

Materials and Methodology:-

Expasy: - To download human protease sequence.

Wormbase database:- To download c.elegan's cysteine protease sequence.

ClustalX: - For multiple sequence alignment.

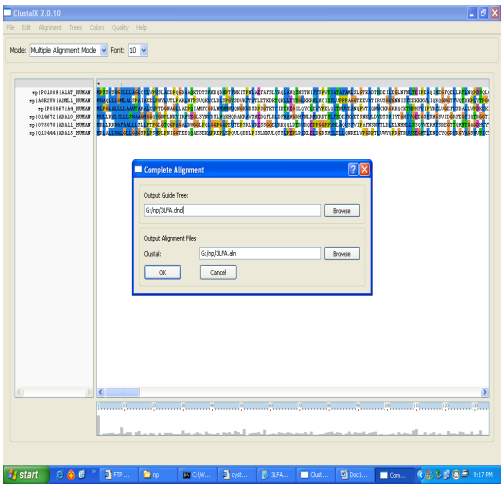

Fig: c.elegan's protein sequences in clustalx

TreeView:- It is used to visualize MSA in graphical form.

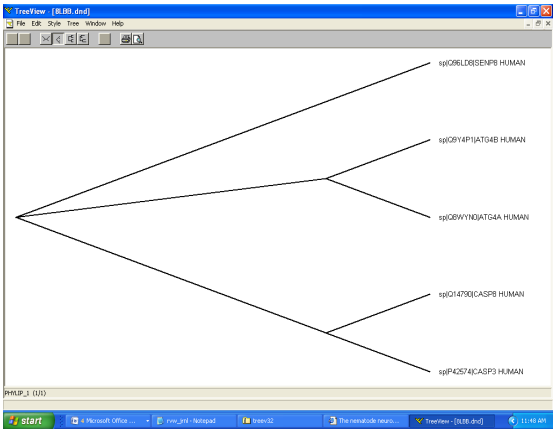

Fig: cladogram

HMMER: - to perform sensitive database searches to identify distant members of sequence families.

Result 1:-

The hmm result is in three steps

1 hmmbuild - build a hidden Markov model from an alignment

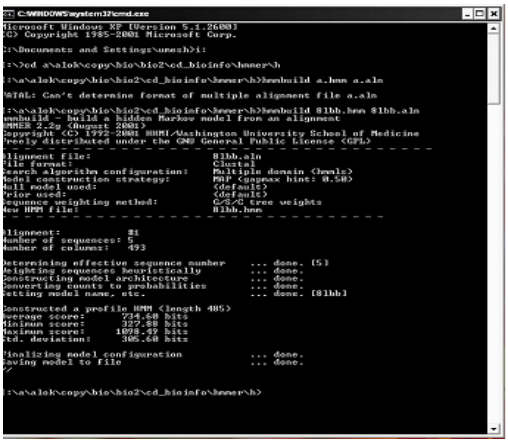

Fig: result of hmmbuild

HMMER 2.2g

Alignment file: 8lbb.aln

File format: Clustal

Search algorithm configuration: Multiple domain (hmmls)

Model construction strategy: MAP (gapmax hint: 0.50)

Null model used: (default)

Prior used: (default)

Sequence weighting method: G/S/C tree weights

New HMM file: 8lbb.hmm

- -

Alignment: #1

Number of sequences: 5

Number of columns: 493

Determining effective sequence number ... done. [5]

Weighting sequences heuristically done.

Constructing model architecture done.

Converting counts to probabilities ... done.

Setting model name, etc. done. [8lbb]

Constructed a profile HMM (length 485)

Average score: 734.60 bits

Minimum score: 327.88 bits

Maximum score: 1098.49 bits

Std. deviation: 305.60 bits

Finalizing model configuration ... done.

Saving model to file ... done.

//

I:\a\alok\copy\bio\bio2\cd_bioinfo\hmmer\h>hmmcalibrate 8lbb.hmm

2 hmmcalibrate -- calibrate HMM search statistics

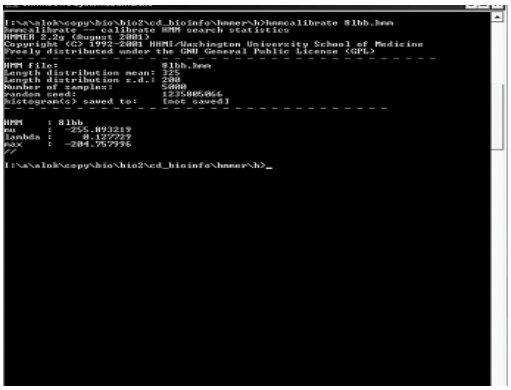

Fig: result of hmm calibrate

hmmcalibrate -- calibrate HMM search statistics

HMM file: 8lbb.hmm

Length distribution mean: 325

Length distribution s.d.: 200

Number of samples: 5000

random seed: 1235805066
HMMER 2.2g (August 2001)

histogram(s) saved to: [not saved]

HMM : 8lbb

mu : -255.893219

lambda : 0.127729

max : -204.757996

//

3 hmmsearch - search a sequence database with a profile HMM

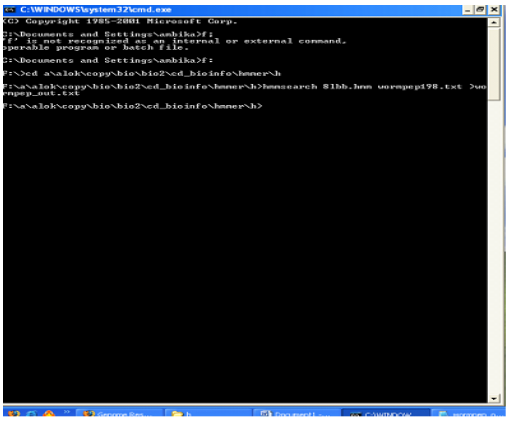

Fig: Result of hmmsearch

Alignments of top-scoring domains:

Y87G2A.3: domain 1 of 1, from 1 to 436: score 124.3, E = *>mDFsRnLYDiGEqLDsEDLAsLKFLsLDYiPqRKqEPmed
+ E L ++ + l++e
Y87G2A.3 1 ----------MTEEIL---------- 20
KQGVGIVETSLTFE-
litlrmleledlpetkELvwiLgrkdlLiTekdkllsdvearLwftyRkk
+ +e + +++ Lg+ + + ++ v +r wftyR+
Y87G2A.3 21 --
PPFCESFERISIDNFPIFALGKEISKEDGIEAMKKYVTSRFWFTYRRD 68
fsAigvmLYqisEgtgpssdrgwgcmLrcgisKCKLDDDmnLLDiqmime
fs+ig gtgps d gwgcmLrc qm +
Y87G2A.3 69 FSPIG-------- 95
GTGPSTDQGWGCMLRCA----------
QMLLG
qavvlrelgrDwkwrkrklipksyvkiLnpFldrkdswssshqigqmgv
+ r+ gr w + y kiL+ F d kd +s+hqi+qmgv
Y87G2A.3 96 EVLLRRHIGRHFEW- 144
DIEKTSEIYEKILQMFFDEKDALYSIHQIAQMGVT

eGkELCGvmsIsdwygpntvAqvlldnsyfddwsslgvcisi nNtnfike
eGk +s w+gpnt Aqv++++ fddws +v+++ +N + ++
Y87G2A.3 145 EGK------ EVSKWFGPNTAAQVMKKLTIFDDWSNIAVH VALDNILVKED 188
isklcrltaliarsgtdlDadnlretFealkdeskqaadsT....wepLl a s Da l ++ + + ++ ++w+pLl
Y87G2A.3 189 ------- AITMATSYPSEDAVKLIMENGLVDKNRLSLS PGNiipeWRPLL 231
lllplrlgldhsnsesfvcslkshgemgiilGtdGqkankaeyfigF kGd
l++plrlgl+ +n +++ ++ + ++ ++ G +G+ +n a yf g G
Y87G2A.3 232 LMIPLRLGLTTIN- PCYLSAIQEFFKIPQCVGIIGGRPNHALYFVG MSGS 280
kLifLdgK.....pkqffiqaceGd.....eldcgIp
kL +Ld++ ++++ ++ + e d++ ++++ + + l+ +p ++++ +Y87G2A.3 281
KLFYLDPHycrpkTESTAKMYAEKDstattddvgfs HLEELVPlpsqtad 330
.....dETdsceqpplrMdiasldpRYIPveadflffckTeadcq nwrn
++ +d T +c q l + dp + ++ fc T + + n +

The output of these files is shown below- Taking the first line of the output above, the meaning of the each of the fields is explained below: Sequence Domain seq-f seq-t hmm-f hmm-t score E-value

-------- ------- ----- ----- ----- ----- ----

Y87G2A.3 1/1 1 436 [. 1 485 [] 124.3 9.3e-34

Domain:1/1 means the first domain of the number of domains detected(1)

Seq-f: sequence from(start)

Seq-t: sequence to (end)

Result 2:-

Histogram of all scores:

Score	obs	exp
(one = represents 6 sequences)		
<-525	4	-|=
-525	1	0|=
-524	0	0|
-523	1	0|=
-522	1	0|=
-521	1	0|=
-520	1	0|=
-519	1	0|=
-518	0	0|
-517	0	0|
-516	1	0|=
-515	2	0|=
-514	0	0|
-513	3	0|=
-512	4	0|=
-511	1	0|=
-510	8	0|==
-509	1	0|=
-508	2	0|=
-507	1	0|=
-506	0	0|
-505	1	0|=
-504	5	0|=
-503	3	0|=
-502	5	0|=
-501	2	0|=
-500	4	0|=
-499	6	0|=
-498	5	0|=
-497	1	0|=
-496	2	0|=
-495	2	0|=
-494	3	0|=
-493	6	0|=
-492	8	0|==
-491	5	0|=

-490 7 0|==

Conclusion:-

At the last of this project, we got the druggable protein domains in *C.elegans* of human counterpart. By using this output we can better understand the signaling pathways (disease process) and be able to inhibit it.

Acknowledgement:-

We would like to express deep sense of gratitude to Mr. Alok Shekhar for his guidance, persuasion, sincere work, valuable suggestion and enthusiasm, which helped us a lot.

References:

1. Wood, William Barry (1988). "Chapter 1: Introduction to *C. elegans* Bioloogy". in Wood, William Barry. *The Nematode* Caenorhabditis elegans. Cold Spring Harbor Laboratory Press. p. 1. ISBN 0-87969-433-5. http://books.google.com/?id=LTEPi6VlZZkC&dq=Caenorhabditis+elegans+length&printsec=frontcover&q=Caenorhabditis%20elegans%20length. Retrieved 2009-12-13.
2. Brenner, S. (May 1974). "The Genetics of *Caenorhabditis elegans*" (PDF). *Genetics* **77**: 71–94. http://dev.wormbase.org/papers/31_Brenner74.pdf.
3. Alberts, B. *et al.* (2008). *Molecular Biology of the Cell*, 5th edition. Chapter 22, page 1321.
4. Nayak, S; Goree, J; Schedl, T (Jan 2004). "*fog-2* and the Evolution of Self-Fertile Hermaphroditism in *Caenorhabditis*". *PLoS Biology* **3** (1): e6. doi:10.1371/journal.pbio.0030006. ISSN 1544-9173. PMID 15630478. PMC 539060. http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pbio.0030006.
5. Kiontke, K, K; Sudhaus, W, W (Jan 2006). "Ecology of Caenorhabditis species." (Free full text). *WormBook : the online review of C. Elegans biology*: 1–14. doi:10.1895/wormbook.1.37.1. PMID 18050464.
6. Shaham S., S (March 14 2006). "Worming into the cell: Viral reproduction in Caenorhabditis elegans" (PDF). *Proc Natl Acad Sci USA* **103** (11): 3955–56. doi:10.1073/pnas.0600779103. ISSN 0027-8424. PMID 16537467. PMC 1449626. http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1449626&blobtype=pdf.
7. Gal TZ, Glazer I, Koltai H (2004). "An LEA group 3 family member is involved in survival of C. elegans during exposure to stress". *FEBS Letters* **577** (1-2): 21–26. doi:10.1016/j.febslet.2004.09.049. PMID 15527756.
8. Sulston JE, Horvitz HR (March 1977). "Post-embryonic cell lineages of the nematode, Caenorhabditis elegans". *Dev. Biol.* **56** (1): 110–56. doi:10.1016/0012-1606(77)90158-0. PMID 838129.
9. Kimble J, Hirsh D (June 1979). "The postembryonic cell lineages of the hermaphrodite and male gonads in Caenorhabditis elegans". *Dev. Biol.* **70** (2): 396–417. doi:10.1016/0012-1606(79)90035-6. PMID 478167.

Site Searches:

1. www.clustal.org/-18k
2. www.expasy.org
3. ftp://ftp.genetics.wustl.edu/pub/eddy/hmmer/2.2g/
4. www.nature.com/uidfinder/10.1038/nrd1552
5. www.ncbi.nlm.nih.gov/pubmed
6. ftp://selab.janelia.org/pub/software/hmmer/current/_user guide.pdb
7. www.treeview.net/tv/download.asp - 6k
8. ftp.wormbase.org
9. http:www.wormbase.org/wiki/index.phs/wsi15g