



Research Journal of Pharmaceutical, Biological and Chemical

Sciences

Membrane Drug Target Identification in Mycoplasma Pneumonia- A Subtractive Genomic Approach.

A Ranganadha Reddy¹, K Praveen Kumar², M Indira¹, TC Venkateswarulu¹, D Ramya Krishna¹, and S Karisma¹, and K Supriya¹.

¹Department of Biotechnology VFSTR University Guntur Andhrapradesh 522 213, India. ²Department of Information Technology VFSTR University Guntur Andhrapradesh 522 213, India.

ABSTRACT

The ease of use of genome sequences of pathogens has provided a remarkable quantity of information that can be used in identification of drug targets and vaccine production. The application of in silico subtractive genomics technology drastically shortens the time required for such purposes. Subtractive genomics approach enables the subtraction dataset between the host and pathogen. The aim of present work is to identify the Drug targets in Mycoplasma Pneumonia by subtractive genomic approach which helps to identify the homolog's and Non homologs between the host and pathogen. The availability of complete Proteome sequence information of Mycoplasma Pneumonia from NCBI-FTP site helps to perform the BLAST against human proteome database. The screening of these proteins using different database like essential genes (DEG), metabolic pathways KEGG helps to identify essential proteins which may be used as drug targets against Mycoplasma pneumonia. Prediction of these essential proteins of the outer membrane of the pathogen also could be potential vaccine targets

Key words: Pneumonia, Subtractive genomic approach, ICM molsoft, NCBI-FTP

*Corresponding author



INTRODUCTION

There are a lot of ways to detect potential drug target such as uncharacterized vital genes, virulence genes, species-specific gene and membrane transporter etc. [1] Comparative genomics provides a new approach to identify new drug targets amongst earlier known targets based on their related biological function in pathogen and host. In the planned work subtractive genomic approach is used, where subtraction dataset comparing two genomes i.e. pathogen and host. This approach is effectively used in quite a few other bacteria such as Pseudomonas aeruginosa[2], Helicobacter pylori[3] etc. The attempt has been made to find the least number of genes necessary for a self-replicating cell, since the whole genome of Mycoplasma has been sequenced. A smallest gene set necessary for a species, which could be deduced from conserved genes in the analyzed genome.[4] "The least possible group of genes that would be adequate to retain a functioning cellular life form under the most favorable conditions possible, i.e. in the presence of full complement of essential nutrients and in the absence of environmental stress" is defined as minimal gene set or vital genes.[5][6][7] In Mycoplasma genitalium 265-350 protein coding genes are identified as vital under laboratory growth condition, which is orthologous to the Mycoplasma pneumonia.[8] In the subsequent work subtractive genomics and Database of Essential Gene (DEG) is used to analyze the genes of Mycoplasma pneumoniae for finding potential target at the outer surface of pathogen, might be used as drug target. Mycoplasma pneumoniae is a potential bacterial pathogen which lacks cell wall and surrounded by a cytoplasmic membrane. It causes atypical pneumonia in human Mycoplasma pneumonia is transmitted from person-to-person contact through respiratory secretions during coughing and sneezing. The incubation period is usually 14-21 days. The whole genome of Mycoplasma pneumoniae has been sequenced. The M129 strain of Mycoplasma pneumoniae is linear single stranded of length 816,394 base pairs with an average G+C contain of 40.0 mol %.[9] All the major classes of cellular process and metabolic pathway are briefly described. A number of activities/functions present in Mycoplasma pneumoniae according to experimental substantiation, but genes or proteins involved in motility, chemotaxis and management of oxidative stress are not known still. [10]

MATERIAL AND METHODS

DATABASES

Database: A database is an assemblage of information that is structured so that it can easily be accessed, managed, and updated. ^[11]

Data mining or discovery: Data mining or knowledge discovery is the computer-oriented process of digging and analyzing large volumes of data and finally extracting the importance of the data. Data mining can be performed on data represented in quantitative, textual, or multimedia forms.^[12]

NCBI (National Center for biotechnology information): The National Center for Biotechnology Information (NCBI)^{[18][13]} is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health. The NCBI houses genome sequencing data in GenBank and an index of biomedical research articles in PubMed Central and PubMed, as well as other information relevant to biotechnology. All these databases are available online through the Entrez search engine.

Swiss-Prot: Swiss-Prot was created in 1986 by Amos Bairoch and developed by the Swiss Institute of Bioinformatics and subsequently developed by Rolf Apweiler at the European Bioinformatics Institute.^{[13][14]} Swiss-Prot aimed to provide reliable protein sequences associated with a high level of annotation a minimal level of redundancy and high level of integration with other databases.

Protein Data Bank: PDB consists of 3D (Three Dimensional) data of experimentally determined structures of proteins and nucleic acids^{[14][15]} established at Brookhaven National Laboratory.^{[15][16]} The archive is managed by the Worldwide Protein Data Bank organization (wwPDB), whose mission is to ensure that a single, global PDB data archive is and will remain freely and publicly available.^{[16][17]}

NCBI (National Center for biotechnology information): The National Center for Biotechnology Information (NCBI)^[18] is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of

2016

RJPBCS



Health. The NCBI houses genome sequencing data in GenBank and an index of biomedical research articles in PubMed Central and PubMed, as well as other information relevant to biotechnology. All these databases are available online through the Entrez search engine.

DEG (Database of Essential Genes): Database of Essential Genes (DEG) that contains all the essential genes currently available. These genes include the essential genes identified in the genomes of *M.genitalium*, *H.influenzae*, *V.cholerae*, *S.aureus*, *E. coli* and *Saccharomyces cerevisiae.*, in which the essential genes are collected from a large number of related references. The essential genes in yeast genome were extracted from the yeast genome database which is maintained by the Munich Information Center for Protein Sequences.

TOOLS

Basic Local Alignment Search Tool **(BLAST)**: Nowadays Similarity searching, including sequence comparison, is one of the principal techniques used by computational biologists and has found widespread use among biologists in general. The most popular tool for this purpose is BLAST ^{[17][18]} (Basic Local Alignment Search Tool) which performs comparisons between pairs of sequences, searching for regions of local similarity. NCBI BLAST is available from the NCBI ^{[18][13]} (National Center for Biotechnology Information).

MOLSOFT: ICM MOLSOFT algorithm was adopted for comparative modeling which provides an accurate and efficient module to build loops and side chains found non-identical in sequence. ICM mol soft algorithm contains robust modeling tools and high levels of accuracy with fast model building.^[19]

Real Space Automated Conformer Generation (RAPPER): RAPPER ^{[20][21]} is an ab initio conformational search algorithm for restraint-based protein modeling. It has been used for all-atom loop modeling, whole protein modeling under limited restraints comparative modeling, ab initio structure prediction, structure validation and experimental structure determination with X-ray and nuclear magnetic resonance spectroscopy. Web interfaces are available on this website for Ramachandran plot analysis.

METHODOLOGY

RETRIEVAL OF TARGET SEQUENCE:

The protein sequence of the protein Dihydroxy acetone kinase system protein of mycoplasma pneumoniae had been retrieved from UniProt, and saved in fasta format that gives the specific information regarding the number of amino acids in the sequence and other sequence related information

PERFORMING TEMPLATE SEARCH:

The protein sequence of the protein Dihydroxy acetone kinase system protein uncharacterized had been retrieved from UniProt and the search for the template had been done using Blast algoritham. Template selection by blast searching Algorithm

HOMOLOGY MODELING USING MOLSOFT:

The query is loaded by going to windows and clicking on sequence editor. Single letter residues are then displayed. Color residues are displayed next.

PDB search is done by going to homology and giving the query. First template in the result is loaded. Then the two sequences are and aligned by freeze in align of homology. The model is build by going to homology and selecting homology model.

Homology modeling

The steps to creating a homology model are as follows:

Identify homologous proteins and determine the extent of their sequence similarity with one another and the unknown.

2016

RJPBCS



- Align the sequences.
- > Identify structurally conserved and structurally variable regions
- Generate coordinates for core (structurally conserved) residues of the unknown structure from those of the known structure(s).
- > Generate conformations for the loops (structurally variable) in the unknown structure.
- Build the side-chain conformations.
- Refine and evaluate the unknown structure.

FLOW CHART OF HOMOLOGY MODELING



Figure-1: The four main steps of homology modeling of protein i.e. template selection, target-template alignment, model building and model quality evaluation ^[22]

VALIDATION OF MODELED PROTEIN USING RAPPER:

As the loop of the structure is built up and the terminals are removed, the modeled structure is now given for validation through RAPPER. The structure is given in .pdb form to it to analyze the structure and to generate the RAMACHANDRAN plot of the structure to find the validity of the structure. This is again also carried out after minimization of the structure is done

RESULTS AND DISCUSSIONS

Retrival of protein sequence from Swissprot :

The details of the Dihydroxy acetone kinase system protein like amino acid lengths of N-and C-terminus, helix lengths etc were obtained from Swiss-Prot database with accession number **P75231.** The details of sequence annotation are shown below.

Blast search against Human Proteome Database:

```
>sp|P75231|Y547_MYCPN Uncharacterized protein MG369 homolog OS=Mycoplasma
pneumoniae (strain ATCC 29342 / M129) GN=MPN_547 PE=3 SV=1
MSTVNLSNFIDMLRLGCKNIANNFEYINQLNVFPVPDGDTGTNMKVTLSEAFKKLESEIS
HIKSFSDLGKSFTRDLLLFSRGNSGVIFSQIMKGFFSDMISTKTATETELGIEDFATAFI
KAEEVAYKNVSKPVEGTMLTVIRLISTDFKNQKNRAKTVQKLFEQVIKTAWQTVKKTPQM
LPVLKASGVVDSGAYGFACFLEGMLSFYGEKATLNDGKLTSAELSQMTISGEKHVTEEEF
GYCTEYVLKLGMSVSQEVEKQKFNQKKFESKVSKIATSVVVASDKDNGFVKVHAHTEKPN
LLLELGLNYGEFELVKIENMNLQVAKQKPAPVKRNIKPAIVVTVPTEAFADRIREDYDIQ
```



_					
	Search	Genomic Biology⊨ Human Genome Resources ►BLAST			
	Map Viewer	Go Clear			
		BLAST Human Sequences. An alternate informative description: feter an accession, gl, or a sequence in FASTA format: WTGFINNLNRAEVHQQLAQLLQVPSIQSLEIIDISFYDKDHVVGAMLRYENGKRMKALSR RYNINDMKSDTNYHADVYRRIISSIQTHKQL#LSDLLIVDGGIAQINTYTKYFASFEN VTGFIIGLANTRRGTENHVUIDINTTINDKNPIFAFALITIQEEVDSFAKHNAFKRVS RARGNPELLQIEGVGRKTVQILLDNFGTNAKHWSCFFERIITVYSC Or, choose a file to upload Browse Set subsequence: (optional)	• <u> </u>		
		From: To:			
		Database: RefSeq protein 38764 sequences			
		Program: BLASTP: Compare protein sequences			
	BLAST Home Re	Basic Local Alignment Search Tool cent Results Saved Strategies Help			My NCBI [Sign In] [Regis
		atting Results - VH4RCVK8016			
		options Download			
	sp P75350 U	VRC_MYCPN UvrABC system protein			
	Descrip	y ID ld[10881 tion sp[P75350[UVRC_MYCPN UVrABC system protein C OS=Mycoplasma pneumoniae GN=uvrC PE=3 SV=1 ype amino acid sgth 586	Des	se Name gp/9606.9558/hs_refp scription Homo sapiens RefSeq protein Program BLASTP 2.2.23+ ▶ <u>Citation</u>	
	🕕 No signifi	cant similarity found. For reasons why, <u>click here</u>			
	Other repo	rts: Þ <u>Search Summarv</u>			

Figure-2: Shows Dihydro acetone kinase protein of Mycoplasma Pneumoniae is non homologous to Human proteome database.

Database of essential Genes Results:

DEG Database of Essential Genes
 Image: State
 Sources
 CHatton
 Version
 Links
 Contact

 Psp: P75350;UVRC_MYCPN
 VerBaC system
 protein c OS-Mycoplasma
 pneumoniae
 GN=uvcC

 MITUNNTLAFELKERAPENCOLVENDUGGEVENUSSENUS
 MITUNNTUAFELKERAPENCOLVENDUGGEVENUSSENUS
 Contact
 GN=uvcC

 MITUNNTUAFELKERAPENCOLVENDUGGEVENUSSENUS
 MITUNNTUAFELKERAPENCOL
 GN=uvcC
 GN=uvcC

 VEFSNRYFGEFASSERFELEDIVDLELKEPELENCAPHENSHUTETINGENGEMEELUNNG
 GN=uvcC
 GN=uvcC
 GN=uvcC

 VEFSNRYFGEFASSERFELEDIVDLELKEPELENCAPHENSHUTETINGENDUGENCOCCOCCED
 GTAFIKAGSHDFISOVYEEKVULAITVFAVUNQISSINGQUFEELFODDEKEVESLUTER
 GTAFIKAGSHDFISOVYEEKVULAITVFAVUNQISSINGQUFEELFODDEKEVESLUTER

 WIGSTINUNARGSTUNINGLISSIGELEIDISFYDKDHVVGAMLERVERKENKALSER
 WIGSTINUNARGSTUNINGLISSIGELEIDISFYDKDHVVGAMLERVERKENKALSER
 VTOFITIGLAKNTRAGTUNUTININGKNESSC

 VTOFITIGLAKNTRAGTUNUTUTINTINIDKNTEFFAULTIGEVDSSC
 GN=uvcS
 GN=uvcS
 GN=uvcS
 Prokaryotes Blast Search Browse Eukaryotes Blast Search Browse Search Clear

DEC	Database of Essential Genes
	Sources Citation Version Links Contact
	BLASTP 2.2.10 [Oct-19-2004]
Prokaryotes	BLASTF 2.2.10 [Oct-19-2004]
Blast	Reference:
Search	Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zheng Neb Miller, and David J. Lipman (1997),
	"Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 2513389-3402.
Browse	
	Query= sp[P75350]UVRC_MYCPN UvrABC system protein C OS=Mycoplasma pneumoniae GN=uvrC PE=3 SV=1
Eukaryotes	(586 letters)
Blast	Database: deg-p.aa 5200 seguences; 1,757,801 total letters
Search	
Browse	Searchingdone
Browse	Score E
	Sequences producing significant alignments: (bits) Value
Click Here if your browser does	12045058_1 DEG10060174 Mycoplasma genitalium G37, uvrC, <u>684</u> 0.0
not support Flash and Script.	15926730_1 DEG10020104 Staphylococcus aureus N315, uvrC, 237 2e-63 15608558 1 DEG1010231 Mycobacterium tuberculosis N37Kv, u 149 9e-37
Prokaryotes	15608558_1 DEG10100231 Mycobacterium tuberculosis H37Rv, u <u>149</u> 9e-37
	16129695_1 DEG10040300 Escherichia coli MG1655, ydjQ/cho, <u>67</u> 6e-12 18497738 1 DEG10120259 Francisella novicida Ull2, FTN 114 29 0.98
Blast	118497025_1 DEG10120106 Francisella novicida U112, FIN_041 29 0.98
Search	
Bearen	>12045058_1 DEG10060174 Mycoplasma genitalium G37, uvrC, excinuclease ABC, C subunt
Browse	Length = 597
	Score = 684 bits (1765), Expect = 0.0
Eukaryotes	
Blast	Query: 4 VNNTLAFKLKNAPHKPGCYLWKDDAGQVLYVGKAKDIFKRVHHYFNPNRSFKTRALVERI 63 + L KLK AP KPGCYLWKD G+VLYVGKA +IF RVH YF N +KT+ L +I
	+ L RLR AF REGULIARD GEVLEVGRA HE RVH IF N HRIFL HI Sbjet: 1 MTTNLKOKKTAPKRGCYLWKDSNGKVLYVGRASNIFNRVHOYFOKNNPYKTOLLSSOI 60
Search	
	Query: 64 ADVEYVILKNENDALNLEAKLIKQYKPRFNLVLKENNGYLYFFITASVKPTLELGRRYEF 123 +DV++ ILK+ENDALNLEAKLI QY+PRFNLVLK+NNGYLYF+IT + KPTLEL R+Y+
Browse	b) + b) + + 1LK+ENDALDLEAKLI QY+FKFNLVLK+NNGYLYF+II + KFILEL KYY Sbjct: 61 SDVDFFILKDENDALNLEAKLINQYQFFNLVLKQNNGYLYFYITKAKFTLELARKYQI 120
	Query: 124 SKNKYFGPFASSKFRLRDIYDLLLKLFPLRKCAPHERGHPCFYYQLKMCMGQCMGEDTPE 183 K FGPFASSKF-LH-HDLLKLFPLRKCAPH++ H+CFY+O+ +CMGCM DT E
k Here if your browser does	Sbjet: 121 KTKCFGPFASSKFKLREIHDLLLKLFFLRKCAPHOKNHPCFYFOMGLCMGCCMOTDIKE 180

Figure-3: Shows Database of essential genes results which indicate Dihydro acetone kinase protein is essential for Mycoplasma Pneumonia

March – April



Template selection by blast searching Algorithm:

Home F	Basic Local Alignment Search Tool Recent Results Saved Strategies Help					<u>My NCBI</u> [Sign In] [R	
NCBI/ BLAST/ bla	astp suite/ Formatting Results - VH4AV5T401S				_		
Edit and Re	esubmit Save Search Strategies Formatting options PDownload						
sp P75350	UVRC_MYCPN UvrABC system protein						
Descr	ery ID c 23117 sp P75350 UVRC_MYCPN UvrABC system protein C OS=Mycoplasma minro acid Length 586		e pdb PDB protein datab BLASTP 2.2.23+ ►				
	ports: >Search Summary [Taxonomy reports] [Distance tree of results] [Related	Structures] [Multiple a	alignment]				
▼ Graphic	Summary						
Show C	Conserved Domains						
	Putative conserved domains have been detect	ted, click on the image b	elow for detailed result	ts.			
1 190 200 300 400 500 500 500							
		and a fill and a					
	Query seq.						
	Query seq. Superfamilies Hulti-donains	uvrC	UvrC_HhH_N				
	Query seq. Superfamilies (IV-YIG superfamily)						
	Query seq. Superfamilies (IN-VIG superfamily) Multi-domains						
equences pro	Query seq. Superfamilies Multi-domains					Lii	
Accession	Query seq. Superfamilies Multi-domains	uvrC	UvrC_HhH_N	superfamily)	▲ <u>E value</u> 3e-19	_	
Accession	Query seq. Superfamilies Hulti-domains roducing significant alignments: Description	uvrC <u>Max score</u>	UvrC_HhH_N Total score	Superfamily		S	
Accession YD6 A C65 A	Query seq. Superfamilies Multi-domains (IVVIG severemily) roducing significant alignments: Description Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease	uvrC <u>Max score</u> <u>93.2</u>	UvrC_HhH_N Total score 93.2	Superfamily) Ouery coverage 16%	3e-19	5	
Accession YD6 A C65 A NRT A	Query seq. If XY15 coveration Multi-donains If XY15 coveration roducing significant alignments: Description Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of Bacillus Stearothermophilus Uvrc 5' E	uvrC <u>Max score</u> <u>93.2</u> <u>75.9</u>	Total score 93.2 75.9	Ouery coverage 16% 34%	3e-19 4e-14	9	
Accession YD6 A C65 A NRT A YCZ A	Ouery seq. Superfamilies Hulti-donains Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Gacillus Stearothermophilus Uvrc S' E Chain A, Crystal Structure Of The C-Terminal Half Of Uvrc >pdbl2N	Max score <u>93.2</u> 75.9 75.9	UvrC_HhH_N Iotal score 93.2 75.9 75.9	Ouery coverage 16% 34% 36%	3e-19 4e-14 4e-14	0 0 0	
Accession YD6 A C65 A NRT A YCZ A YD3 A	Buery seq. Differential moduling significant alignments: Interferential moduling significant alignments: Description Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The C-Terminal Half of Uvrc >rdb12N Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Half of Uvrc >rdb12N	<u>Wax score</u> <u>93.2</u> <u>75.9</u> <u>65.9</u>	UurC_HhH_N Total score 93.2 75.9 75.9 65.9	Superfamily Query coverage 16% 34% 34% 36% 14%	3e-19 4e-14 4e-14 4e-11		
Accession YD6 A C65 A NRT A YCZ A YD3 A YD4 A	Ouery seq. Multi-donains INVEX.secretarily	<u>wvrC</u> <u>93.2</u> <u>75.9</u> <u>65.9</u> <u>64.7</u>	UVCC_HAH_N Total score 93.2 75.9 75.9 65.9 64.7	superfamily <u> Suery coverage</u> 16% 34% 36% 14% 14%	3e-19 4e-14 4e-14 4e-11 1e-10		
Accession YD6_A C65_A NRT_A YCZ_A YD3_A YD3_A YD4_A YD2_A	Ouery seq. Superfamilies Hulti-donains Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The C-Terminal Half Of Uvrc >rdb12N Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease	Max score 93.2 75.9 75.9 65.9 65.9 64.7 64.7	UVPC_HRH_N 93.2 75.9 75.9 65.9 64.7 64.7	Superfamily 0uery coverage 16% 34% 36% 14% 14% 14%	3e-19 4e-14 4e-14 4e-11 1e-10 1e-10	0 0 0 0 0 0 0	
Accession YD6_A C65_A NRT_A YCZ_A YD3_A YD3_A YD4_A YD2_A NRR_A	Buery seq. Effective conversity multi-donains Effective conversity moducing significant alignments: Description Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease	Max score 93.2 75.9 65.9 64.7 64.7 64.7	UVPC_HHH_N 93.2 75.9 75.9 65.9 64.7 64.7	Superfamily Ouery coverage 16% 34% 36% 14% 14% 14% 14%	3e-19 4e-14 4e-14 4e-11 1e-10 1e-10 1e-10	8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	
Accession YD6 A C65 A NRT A YCZ A YD3 A YD4 A YD4 A YD2 A NRR A YD5 A	Buerg seq. Description introducting significant alignments: Introducting significant alignments: n Description Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease	Max score 93.2 75.9 65.7 64.7 64.7 64.7 64.3 63.2	UVPC_HNH_N 93.2 75.9 75.9 64.7 64.7 64.7 64.7 64.3	Superfamily 2007 Coverage 16% 34% 36% 14% 14% 14% 14% 28%	3e-19 4e-14 4e-14 4e-11 1e-10 1e-10 1e-10 1e-10	8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	
Accession YD6 A C65 A NRT A YCZ A YD3 A YD4 A YD2 A NRR A YD5 A OOX G	Buery seq. multi-donains itivitis cournetis moducing significant alignments: moducing significant alignments: moducing significant alignments: Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Civ-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Civ-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure Of The Giv-Yig N-Terminal Endonuclease	Max score <u>33.2</u> <u>75.9</u> <u>65.9</u> <u>64.7</u> <u>64.7</u> <u>64.7</u> <u>64.7</u> <u>64.7</u> <u>64.7</u> <u>64.7</u> <u>64.3</u> <u>29.3</u>	Iotal score 93.2 75.9 65.9 64.7 64.7 64.7 64.3 63.2	Superfamily Ouery coverage 16% 34% 36% 14% 14% 14% 28% 14%	3e-19 4e-14 4e-14 1e-10 1e-10 1e-10 1e-10 3e-10	0 0 0 0 0 0 0 0 0 0	
	Buerg seq. Description introducting significant alignments: Introducting significant alignments: Introducting significant alignments: Description Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease Chain A, Crystal Structure of The Giv-Yig N-Terminal Endonuclease	Max score 93.2 75.9 65.7 64.7 64.7 64.7 64.3 63.2	UVPC_HhH_N 93.2 75.9 65.9 64.7 64.7 64.3 63.2 29.3	Superfamily Duery coverage 16% 34% 36% 14% 14% 14% 28% 14% 12%	3e-19 4e-14 4e-14 1e-10 1e-10 1e-10 1e-10 3e-10 5.0		

Figure-4 : BLAST result for Dihydro acetone kinase protein

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between protein or nucleotide sequences. The program compares nucleotide or protein sequences to sequence in a database and calculates the statistical significance of the matches. The target sequence i.e., Dihydro acetone kinase protein (UniProt ID: P75231) was searched against the protein database by using BLAST tool. From the BLAST results as shown in Figure-2, Figure-3 and Figure-4. we observed that four proteins (PDB IDs: 1BLS-A, 3C65-A, 2NRT-A, 1YCZ-A) are showing the maximum identity with the target sequence. Among the four proteins obtained, 1BLS-A selected for further proceedings.

Three-Dimensional Structure Prediction of Dihydroxy acetone kinase Protein by MOLSOFT Software: Three dimensitional structure of Dihydroxy acetone kinase protein (UniProt ID: P75231) as shown in Figure-5 and Figure-6 was predicted by using the tool MOLSOFT ICM by taking **1BLS-A** as template which was obtained through BLAST results by taking **P75231** as query sequence and performing Protein BLAST against the protein sequence database.

Assessment of the Homology Model of Dihydroxy acetone kinase system protein: The validation of the final model was carried out using Ramachandran plot computed with PROCHECK, program by checking the detailed residue-by-residue stereo-chemical quality of a protein structure. The PROCHECK is used for stereo chemical assessment of the model. The criteria for analysis of stereochemistry of the model includes,

- 1) Main chain conformation in acceptable regions of the Ramachandran plot.
- 2) Planar peptide bonds.
- 3) Side chain conformations that correspond to those in rotamer library.
- 4) Hydrogen bonding of polar atoms if they are buried.
- 5) No bad atom-atom contacts.
- 6) No holes inside the structure.

Ramachandran Plot analysis by RAPPER

A Ramachandran plot (also known as a Ramachandran map or a Ramachandran diagram or a $[\Phi,\Psi]$ plot), developed by Gopalasamudram Narayana Ramachandran and Viswanathan Sasisekharan is a way to visualize dihedral angles Ψ and Φ of amino acid residues in protein structure. It shows the possible conformations of Φ and Ψ angles for a polypeptide. Hence, Ramachandran plot is a useful way of assessing the stereo chemical quality of a protein structure. The final refined model was further assessed by RAPPER server (ramachandran plot analysis), and the results show that Dihydroxy acetone kinase protein had 295 (96.3%)

March – April

2016

RJPBCS

7(2) Page No. 688



residues in favoured region against (~98.0% expected), 35 (5.2%) residues in allowed region against (~2.0% expected) and 12 (3.5%) residues in outlier region as shown in **Figure-7** which shows that the final model is reliable.

Template structure (1BLS-A):



Figure-5:3d Structure Prediction of Dihydroxy acetone kinase system protein C by MOLSOFT:



Figure-6: Shows Query (Dihydroxy acetone kinase) structure was modeled by using ICM molsoft by taking template as 1BLS, which is showing high similarity with query.

Ramachandran plot analysis



March – April





Figure-7: Protein validation study by RAPPER Server

Number of residues in favoured region (~98.0% expected): 295 (96.3%) Number of residues in allowed region (~2.0% expected) : 35 (5.2%) Number of residues in outlier region : 12 (3.5%)

CONCLUSIONS

The large scale genome sequencing projects have increased the availability of completely sequenced genomic and proteomic data in public domain. Use of the DEG database is more efficient than conventional methods for identification of essential genes and proteins and facilitates the exploratory identification of the most relevant drug targets in the pathogen.

The present study has thus led to the identification of several proteins that can be targeted for effective drug design and vaccine development against Mycoplasma pneumonia drugs developed against these will be specific to the pathogen, and therefore less or not toxic to the host. In the present study the numbers of essential gene products (Proteins) 26 are identified in Mycoplasma pneumoniae, by subtractive genomic approach. However these can be further increased, characterized and their role in the survival of the bacteria can be verified. Homology modeling of these targets will help identify the best possible sites that can be targeted for drug design by simulation modeling. Virtual screening against these novel targets might be useful in drug discovery.

In our study we have taken drug target as membrane protein Dihydroxy acetone kinase .The structure of Dihydroxy acetone kinase protein was modeled by taking template as 1BLS in MOLSOFT and model verification done by RAPPER.

REFERENCES

- [1] Galperin MY, Koonin EV Searching for drug targets in microbial genomes. Curr Opin Biotechnol 10,1999, 571-578.
- [2] Sakharkar KR, Sakharkar MK, Chow VT A novel genomics approach for the identifi cation of drug targets in pathogens, with special reference to Pseudomonas aeruginosa. In Silico Biology, 4,2004,355-360.
- [3] Dutta A, Singh SK, Ghosh P, Mukherjee R, Mitter S, et al. In silico identifi cation of potential therapeutic targets in the human pathogen Helicobacter pylori. In Silico Biol 6, 2006,43-47.



- [4] Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N The use of gene clusters to infer functional coupling. Proc Natl Acad Sci USA, 96, 1999, 2896- 2901.
- [5] Koonin EV How many genes can make a cell: the minimal-gene-set concept. Annu Rev Genomics Human Genetics 1, 2000, 99-116.
- [6] Koonin EV Comparative genomics, minimal gene-sets and the last universal common ancestor. Nat Rev Microbiology ,1, 2003, 127-136.
- [7] Gil R, Silva FJ, Peretó J, Moya A, Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev 68,2004, 518-537.
- [8] Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, et al., Global transposon mutagenesis and a minimal Mycoplasma genome. Science 286,1999, 2165-2169.
- [9] Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, et al., Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumonia*, Nucleic Acids Res 24,1996, 4420-4449.
- [10] Wenzel R, Herrmann R, Cloning of the complete Mycoplasma pneumoniae genome. Nucleic Acids Res 17,1993, 7029-43.
- [11] K.Praveen Kumar, U.Janardhan Reddy, A.Ranganadha Reddy, N.ShyamalaDevi ,Homology modeling and Docking Studies of MPT 51 Protein in Mycobacterial leprae International Journal of Chemtech Research,6(7),2014, 3864-3882.
- [12] Venkateswarulu TC, Kodali VP, Ranganadha Reddy A,John DB,Rahul RN,Indira M,Narayana AV,Comparative Studies of H5N1 Gene Segments with other Subtypes of Influenza- A Virus, Research Journal of Pharmaceutical, Biological and Chemical Sciences. 5, 2014, 1417–1429.
- [13] Ranganadha Reddy A, Venkateswarulu TC, JohnBabu D, M.Indira, Homology modeling studies of human genome receptor using modeller, swiss-model server and esypred-3d tools, International Journal of Pharmaceutical Sciences Review Research vol 30(5), 2015,1-6.
- [14] .Séverine Altairac, "Naissance d'une banque de données: Interview du prof. Amos Bairoch". Protéines à la Une, August 2006. ISSN 1660-9824.
- [15] Berman, H. Acta Crystallogr A: Foundations of Crystallography 2008, Vol 64,:pp88–95.
- Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. Journal of Molecular Biology 1977, 112,:pp535–542.
- [17] Berman, H. M.; Henrick, K.; Nakamura, H. Nat Struct Biol 2003, 10, 980.
- [18] Ranganadha Reddy A, Sreedhara.R.Voleti, Lakshmi Padma Ch,Towards the prediction f a threedimenional structure of human glp-1receptor. Research Journal of Pharmaceutical, Biological and Chemical Sciences, 4, 2013, 151-168.
- [19] Ranganadha Reddy A, Venkateswarulu TC, JohnBabu D, M.Indira, Homology modeling studies of human genome receptor using modeller, swiss-model server and esypred-3d tools, International Journal of Pharmaceutical Sciences Review Research vol 30(5), 2015,1-6.
- [20] Schapira M., Raaka B.M., Samuels H.H. and Abagyan R. In silico discovery of novel Retinoic Acid Receptor agonist structures BMC Structural Biology,2001, 1:1.
- [21] Ranganadha Reddy A Salamuddin SK, Venkata Rami Reddy Ch, Homology modeling studies of ras protein in cancer Research Journal Pharmaceutical, Biological and Chemical Sciences, 5,(5) ,2014,273-283
- [22] Ranganadha Reddy A, Venkateswarulu TC, Indira M, Narayana AV, Naga Lohita T, Sriharsha M, Identification of Membrane Drug Targets by Subtractive Genomic Approach in Mycoplasma Pneumonia, Research Journal of Pharmacy and Technology, 8(9), 2015, 1209-1216.