Journal of Antivirals & Antiretrovirals - Open Access

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jaa.1000020

JAA/Vol.2 Issue 2

Definition of Potential Targets in *Mycoplasma pneumoniae* Through Subtractive Genome Analysis

Gupta Sunil Kumar[1, 2*], Singh Sarita[1, 2], Gupta Manish Kumar[3], Pant KK[2] and Seth PK[1]

[1]*Bioinformatics Centre, Biotech Park, Sector-G, Jankipuram, Lucknow-226021, Uttar Pradesh, India*
[2]*Department of Pharmacology & Therapeutics, Chhatrapati Shahuji Maharaj Medical University, Chowk Lucknow-226003, Uttar Pradesh, India*
[3]*Department of Bioinformatics, University Institute of Engineering and technology, Chhatrapati Shahuji Maharaj University, Kanpur, India*

Abstract

Whole genome sequencing technology provided expensive information for the identification of new therapeutic targets in pathogens over and above human genome. Subtractive genomic approach is extremely informative technique to identify the potential targets, which are expected to be essential genes or proteins in pathogen but absent in host, which can be used as drug target. Besides it uncharacterized proteins, which are present on the exposed surface of pathogen, may also be consider as drug target. In present study subtractive genomic approach has been used to identify therapeutic target in *Mycoplasma pneumoniae*, which is atypical pneumonia causing pathogen in human. The subsequent analysis revealed that 732 genes were coding 693 proteins in *M. pneumoniae* out of which 71 proteins were duplicate and 220 proteins were found essential nonhuman homolog. Further analysis of these non human homologous proteins predicted that 27 essential proteins were involved in unique metabolic pathways of *Mycoplasma pneumoniae*. Therefore these 27 essential proteins may serve as therapeutics target. Protein localization predictions of 220 essential proteins were exposing that 12 proteins were present on the exposed surface of pathogen. These exposed surface proteins could be the possible drug targets as well.

Keywords: Therapeutic target; Homologs; Chemotaxis; Orthologous; *Mycoplasma pneumoniae*

Introduction

As of December 2009, the complete genome sequence was known of about 2274 viruses (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html), 1007 bacterial species and roughly 56 eukaryote organisms, of which about half are fungi (http://www.ncbi.nlm.nih.gov/genomeprj) and a number of bioinformatics tools are also developed to analyze those genome (Kaminski, 2000). Completion of Human Genome Project is one of the major revolutions in the field of drug discovery against human pathogen. At present time genomic approach is in tradition (Galperin and Koonin, 1999). Identification of novel therapeutic targets is one of the major tasks in order to design a novel drug.

There are many approaches to identify potential drug target such as virulence genes, uncharacterized essential genes, species–specific gene, unique enzyme and membrane transporter etc (Galperin and Koonin, 1999). Comparative genomic provide a new approach to identified novel drug target among previously known targets based on their related biological function in pathogen and host.

In the proposed work subtractive genomic approach is used, where subtraction dataset comparing two genomes i.e. pathogen and human. This approach is successfully used in many other bacteria such as *Pseudomonas aeruginosa* (Sakharkar et al., 2004), *Helicobacter pylori* (Dutta et al., 2006), *Burkholderia pseudomalleii* (Chong et al., 2006) etc.

The effort has been made to find the minimal number of genes required for a self-replicating cell, since the complete genome of *Mycoplasma* has been sequenced. A minimal gene set required for a species, which could be deduced from conserved genes in the analyzed genome (Overbeek et al., 1999). "A smallest possible group of genes that would be sufficient to sustain a functioning cellular life form under the most favorable conditions imaginable, that is, in the presence of full complement of essential nutrients and in the absence of environmental stress" is defined as minimal gene set or essential genes (Koonin, 2000; Koonin, 2003; Gil et al., 2004). In *Mycoplasma genitalium* 265-350 protein coding genes are identified as essential under laboratory growth condition, which is orthologous to the *Mycoplasma pneumoniae* (Hutchison et al., 1999).

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 cell wall less bacterial pathogen and surrounded by a cytoplasmic membrane only. It causes a typical pneumonia in human (Chanock et al., 1963). *Mycoplasma pneumoniae* is transmitted from person-to-person contact through respiratory secretions during coughing and sneezing. The incubation period is usually 14-21 days. The entire genome of *Mycoplasma pneumoniae* has been sequence. 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 % (Himmelreich et al., 1996).

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 evidence, but genes or proteins involved in motility, chemotaxis and management of oxidative stress are not known still. The M129 strain of *Mycoplasma pneumoniae* is used here because it involves in cytadherence and pathogenicity studies (Wenzel and Herrmann, 1989).

***Corresponding author:** Sunil Kumar Gupta, Bioinformatics Centre, Biotech Park, Sector-G, Jankipuram, Lucknow-226021, Uttar Pradesh, India, Fax: +91 522 4012081; Tel: +91 522 4053010; E-mail: skgupta.res@gmail.com

Received March 15, 2010; **Accepted** April 26, 2010; **Published** April 26, 2010

Citation: Gupta SK, Singh S, Gupta MK, Pant KK, Seth PK (2010) Definition of Potential Targets in *Mycoplasma Pneumoniae* Through Subtractive Genome Analysis. J Antivir Antiretrovir 2: 038-041. doi:10.4172/jaa.1000020

Methodology

Sequence retrieval of host and pathogen

The complete genome, genes and protein sequences of *Mycoplasma pneumoniae* strain M129 as well as *Homo sapiens* were retrieved from the NCBI (National Center for Biotechnology Information) and Swiss-Prot Protein knowledgebase (http://www.expasy.ch/sprot/). From the complete genome sequence data, the genes of the organism that coded for proteins whose sequence were greater than 100 amino acids were selected out. This was on the assumption that proteins less than 100 amino acids in length were unlikely to represent essential proteins, yet be unique to the organism.

Identification of duplicate protein

The *Mycoplasma pneumoniae* proteins were eliminated at 60% using CD-HIT suite (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit) to identify the paralogs or duplicates proteins within the proteome of *Mycoplasma pneumoniae*. The prologs were excluded and the remaining sets of protein were used for further analysis.

Similarity search

The nonparalogs proteins were subjected to NCBI BlastP (http://www.ncbi.nim.nih.gov/blast) (Altschul et al., 1990) against Homo sapiens protein sequences using threshold expectation value 10^{-3} as parameter to find out the nonhuman homologues proteins of *Mycoplasma pneumoniae*. The human homologous were excluded and the list of non-homologs was compiled. The selected nonhuman homologues proteins were then subjected to similarity search using standard NCBI TBLASTN against the Database of Essential Genes (DEG) (http://tubic.tju.edu.cu/deg1). A random expectation value (E-value) cut-off of 10^{-100} and a minimum bit-score cut-off of 100 were used to screen out proteins that appeared to represent essential proteins.

Metabolic pathway analysis

Metabolic pathway analysis of the essential proteins of *Mycoplasma pneumoniae* was done by KAAS server at KEGG (http://www.genome.jp/tools/kaas/) for the identification of potential targets. KAAS (KEGG Automatic Annotation Server) provides functional annotation of genes by BLAST comparisons against the manually curated KEGG GENES database. The result contains KO (KEGG Orthology) assignments and automatically generated KEGG pathways.

Surface protein identification

Prediction of protein localization is an important to predict the protein function and genome annotation, and it can assist the identification of targets. Sub-cellular localization analysis of the essential protein sequences has been done by Proteome Analyst Specialized Subcellular Localization Server v2.5 (PA-SUB) (http://webdocs.cs.ualberta.ca/~bioinfo/PA/Sub/) to identify the surface membrane proteins which could be feasible vaccine target.

Classify functions for the uncharacterized essential proteins

Functional family prediction of the putative uncharacterized essential proteins was done by using the SVMProt web server (http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi.) (Cai et al., 2003). SVMProt utilizes Support Vector Machine for classification of a protein into functional family from its primary sequence.

Result and Discussion

The increasing number of complete bacterial genomes available in the public databases offers new opportunities for understanding the relationship between genotype and phenotype using in-silico genome comparisons. Subtractive genome analysis is an attempt to link genome content and phenotypic features according to the presence or absence of genes. The method is based on the assumption that the genes responsible for a specific function are conserved during evolution but lost in those genomes not showing that phenotype. Therefore, this method is used to search for those genes which are present in a group of genomes having a common phenotype, but which are absent in another group not showing this phenotype, as for instance the capacity to grow in the presence of an antibiotic or the ability to synthesize an outer membrane. This strategy may be a first step in the understanding of adaptive mechanisms of micro-organisms. Although experimental and computational methods have been previously employed for the study of essential genes, to our knowledge, this is the initial report of essential gene or protein identification as probable drug targets in *M. pneumoniae* by subtractive genomics approach. By applying this approach, following results were obtained as described below (Table 1). The objective of that analysis was to find out the essential proteins, which play a key role in survival of bacteria within human and identify them as drug target to block the bacterial pathogenesis.

In silico subtractive/differential genome analysis is a powerful approach for identifying genus- or species-specific genes, or groups of genes that are responsible for a unique phenotype. By this method, one searches for genes present in one group of bacteria and absent in another group. In current study, non-human homolog essential genes of *Mycoplasma pneumoniae* as well as their protein products was identified by applying subtractive genomic approach, which are likely to lead development of drugs that strongly bind with the pathogen. Flow diagram of step by step approach used in current study is described in Figure 1. The above analysis reveals that 693 proteins are present in *Mycoplasma pneumoniae* M129 strain. The duplicate proteins were identified using 60% identity as threshold via CD-HIT tool. Out of 693 proteins 71 were found duplicate or paralogs proteins. Thereafter paralogs were excluded and remaining 590 were underwent for similarity search using BlasatP against human proteome, which resulted 375 proteins were non-human homolog. Among these 375 proteins, 220 proteins were essential proteins of *Mycoplasma pneumoniae*.

The result of metabolic pathway analysis using by KAAS server at KEGG revels that out of these 220 proteins of *Mycoplasma pneumoniae*

Total Number of proteins	693
Protein >100 amino acid	661
Duplicates (>60% identical) in CD-HIT	71
Non-paralogs	590
Non-human homologous proteins (E-value 10^{-3})	375
Essential protein in DEG (E-value 10^{-10})	220
Essential proteins involved in metabolic pathways	112
Proteins involved in unique pathways	27
Membrane associated non-human homolog of essential genes (Outer membrane/Extra-cellular)	12

Table1: Subtractive proteomic and metabolic pathway analysis result for *Mycoplasma pneumoniae*.

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jaa.1000020
JAA/Vol.2 Issue 2

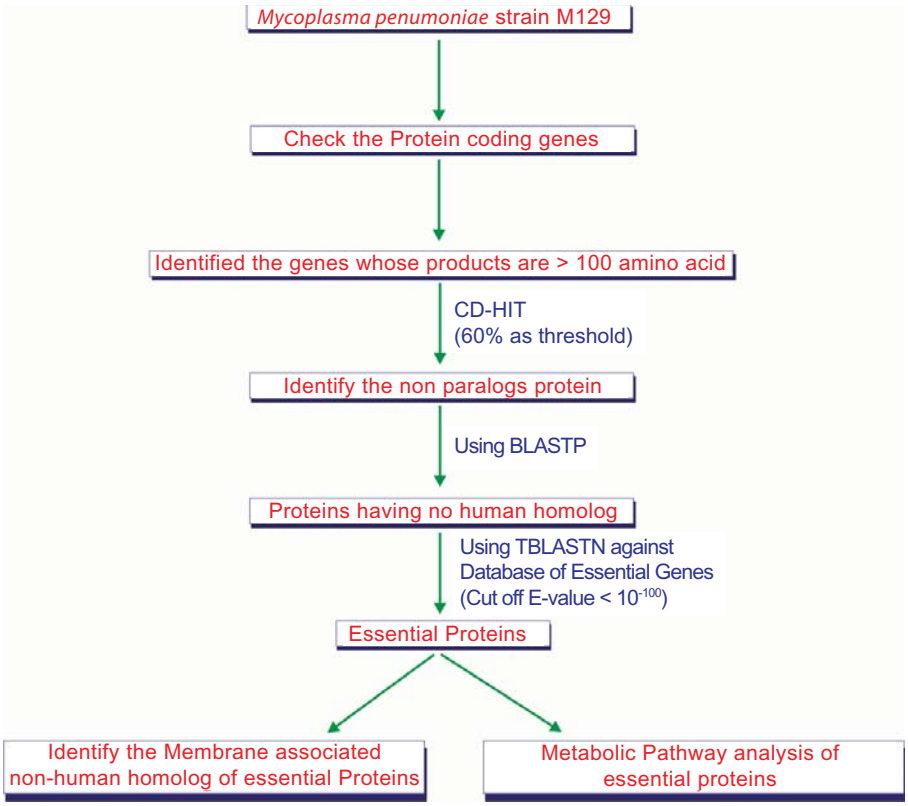

Figure 1: Flowchart of brief Methodology.

S. N.	Protein name	Accession no.	Sub-cellular location
1.	MYCPN Uncharacterized protein MG075 homolog	P75556	**Outer membrane**
2.	MYCPN ATP synthase subunit b	Q50327	**Outer membrane**
3.	MYCPN Uncharacterized protein MPN_438	P75340	**Outer membrane/Extra-cellular**
4.	MYCPN Uncharacterized protein MG144 homolog	P75588	**Extra-cellular**
5.	MYCPN Uncharacterized lipoprotein MG045 homolog	P75056	**Extra-cellular**
6.	MYCPN Uncharacterized lipoprotein MG186 homolog	P75265	**Extra-cellular**
7.	MYCPN Putative adhesin P1-like protein MPN_286	P75491	**Extra-cellular**
8.	MYCPN Uncharacterized protein MPN_586	P75194	**Extra-cellular**
9.	MYCPN Uncharacterized lipoprotein MPN_582	P75198	**Extra-cellular**
10.	MYCPN Uncharacterized lipoprotein MPN_585	P75195	**Extra-cellular**
11.	MYCPN Uncharacterized protein MPN_591	Q50336	**Extra-cellular**
12.	MYCPN Oligoendopeptidase F homolog	P54125	**Extra-cellular**

Table3: Membrane associated protein in *Mycoplasma pneumoniae* with their swiss-prot Accession number and sub cellular locations.

112 essential proteins might be concluded to be unique and are consistently linked with essential metabolic and signal transduction pathways. Testing of drug like molecule against such protein target might be helpful to block the pathogenesis. Metabolic pathway analysis of these 112 essential proteins resulted that 15 proteins are involved in Carbohydrate Metabolism, 9 in Energy Metabolism, 17 in Nucleotide Metabolism, 4 in Amino Acid Metabolism, 4 in Metabolism of Co-factors and Vitamins, 42 in genetic information processing and 21 in environmental information processing (Table 2(Included as Supplement material)). Out of 42 proteins involved in genetic information processing 9 proteins were take part in replication of *M. pneumoniae*. If these proteins are bind by inhibitors then the replication of bacteria will be interrupt. Therefore the functionality of these proteins is needed for replication and pathogenesis. Thus these proteins are important targets for drug development against *M. pneumoniae* infection.

Comparative analysis of the metabolic pathways of the host (Homo sapiens) and the pathogen (*Mycoplasma pneumoniae*) by using Kyoto Encyclopedia of Genes and Genomes (KEGG) reveals 27 proteins were Proteins involved in unique pathways of *Mycoplasma pneumoniae*.

However these 27 unique proteins involved in various metabolic pathway of Mycoplasma pneumonia, which are essential for survival of bacteria in minimal medium as well as their regulatory function. Hence these unique proteins might be good target for drug.

Prediction of sub-cellular location of 220 essential protein of *Mycoplasma pneumoniae* using PA-SUB was result out that 12 proteins were found on the exposed surface of pathogen (Table 4). Out of these 12 proteins most were uncharacterized protein. The functional classification of the 12 putative uncharacterized essential proteins, exposed on surface of pathogen, was performed by using the SVMProt web server based on P value, which is the expected classification accuracy in terms of percentage. 2 proteins were

S. No.	Accession No.	Protein Family Name	R-Value	P-Value (%)
1.	P75556	Transmembrane	6.0	99.0
2.	P75588	Transmembrane	6.0	99.0
3.	Q50327	Hydrolases	6.0	99.0
4.		Sodium-binding	6.0	99.0
5.		Transmembrane	3.4	96.1
6.	P75056	Lipid-binding protein	4.0	97.7
7.	P75340	Lipid-binding protein	2.1	85.4
8.	P75194	Lipid-binding protein	4.7	98.5
9.		Zinc-binding	4.1	97.8
10.		Transferases - Glycosyltransferases	3.2	95.2
11.		Electrochemical Potential-driven transporters	1.8	80.4
12.	P75198	Lipid-binding protein	6.7	99.1
13.		Hydrolases	2.5	90.3
14.	P75195	Lipid-binding protein	1.8	80.4
15.	Q50336	Hydrolases	2.8	92.9
16.	P54125	Zinc-binding	6.0	99.0
17.		Hydrolases (acting on peptide bonds)	6.0	99.0
18.		Hydrolases (Acting on Ester Bonds)	1.9	82.2

Table4: Membrane associated protein and their functions in *Mycoplasma pneumonia*.

classified as transmembrane proteins, 2 as zinc binding, 5 as lipid-binding, 2 as Hydrolases, 1 as Outer membrane (Table 4). Thus these membranes or surface associated non-human homolog proteins of *Mycoplasma pneumoniae* may be used as therapeutic target for vaccine designing.

In the whole study two parallel ways were used to identify the suitable drug target for *Mycoplasma pneumonia*e using subtraction of genomic information. This approach was already successfully used in many organisms such as *Pseudomonas aeruginosa, Helicobacter pylori, Burkholderia pseudomallei, Mycobacterium tuberculosis H37Rv, Salmonella typhi and Neisseria meningitides serogroup B* for drug target identification, which results constructive thoughts for further drug development.

Conclusion

A number of approaches for new vaccine development exist, such as sub-unit protein and DNA vaccines, recombinant vaccines, auxotrophic organisms to deliver genes and so on. Testing such candidates is tedious and expensive. *In-silico* approaches enable us to reduce substantially the number of such candidates to test and speed up drug discovery with least toxicity. The use of DEG database is more efficient than conventional methods for identification of essential genes and facilitates the exploratory identification of the most relevant drug targets in the pathogen. The subtractive genomic approach has been applied in the present study for the identification of several proteins that can be targeted for effective drug design and vaccine development against *M. pneumoniae*. The drugs developed against these will be specific to the pathogen, and therefore less or non toxic to the host. Structural 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 the discovery of novel therapeutic compounds against *M. pneumoniae*.

Acknowledgments

The support of Department of Biotechnology, Ministry of Science and Technology, Government of India, to Bioinformatics Centre at Biotech Park Lucknow is gratefully acknowledged. Also extremely acknowledged to Department of Pharmacology, C.S.M.M. University, Lucknow, U. P. India.

References

1. Chong CE, Lim BS, Nathan S, Mohamed R (2006) In silico analysis of Burkholderia pseudomallei genome sequence for potential drug targets. In Silico Biol 6: 341-346. » CrossRef » PubMed » Google Scholar
2. Dutta A, Singh SK, Ghosh P, Mukherjee R, Mitter S, et al. (2006) In silico identification of potential therapeutic targets in the human pathogen Helicobacter pylori. In Silico Biol 6: 43-47. » CrossRef » PubMed » Google Scholar
3. Galperin MY, Koonin EV (1999) Searching for drug targets in microbial genomes. Curr Opin Biotechnol 10: 571-578. » CrossRef » PubMed » Google Scholar
4. Gil R, Silva FJ, Peretó J, Moya A (2004) Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev 68: 518-537. » CrossRef » PubMed » Google Scholar
5. Hilbert H, Himmelreich R, Plagens H, Herrmann R (1996) Sequence analysis of 56 kb from the genome of the bacterium Mycoplasma pneumoniae comprising the dnaA region, the atp operon and a cluster of ribosomal protein genes. Nucleic Acids Res 24: 628-639. » CrossRef » PubMed » Google Scholar
6. Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, et al. (1996) Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae. Nucleic Acids Res 24: 4420-4449. » CrossRef » PubMed » Google Scholar
7. Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, et al. (1999) Global transposon mutagenesis and a minimal Mycoplasma genome. Science 286: 2165-2169. » CrossRef » PubMed » Google Scholar
8. Kaminski N (2000) Bioinformatics. A user's perspective. Am J Respir Cell Mol Biol 23: 705-711. » CrossRef » PubMed » Google Scholar
9. Koonin EV (2000) How many genes can make a cell: the minimal-gene-set concept. Annu Rev Genomics Hum Genet 1: 99-116. » CrossRef » PubMed » Google Scholar
10. Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. Nat Rev Microbiol 1: 127-136. » CrossRef » PubMed » Google Scholar
11. Lu Z, Szafron D, Greiner R, Lu P, Wishart DS, et al. (2004) Predicting Subcellular Localization of Proteins using Machine-Learned Classifiers. Bioinformatics 20: 547-556. » CrossRef » PubMed » Google Scholar
12. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N (1999) The use of gene clusters to infer functional coupling. Proc Natl Acad Sci USA 96: 2896-2901. » CrossRef » PubMed » Google Scholar
13. Sakharkar KR, Sakharkar MK, Chow VT (2004) A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa. In Silico Biol 4: 355-360. » CrossRef » PubMed » Google Scholar
14. Wenzel R, Herrmann R (1989) Cloning of the complete Mycoplasma pneumoniae genome. Nucleic Acids Res 17: 7029-43. » CrossRef » PubMed » Google Scholar