In silico analyses of metabolic pathway and protein interaction network for identification of next gen therapeutic targets in Chlamydomphila pneumoniae

Bilachi S Ravindranath*¹, Venkatappa Krishnamurthy¹, Venkatarangaiah Krishna² & Kahale Bommaiah Lingaiah Vasudevanayaka³

¹Department of Biotechnology, PES Institute of Technology (Recognized research centre of Kuvempu University), BSK III Stage, Bangalore - 560085, India; ²Department of Biotechnology and Bioinformatics, Kuvempu University, Shankarghatta -577451, India; ³Bapuji Institute of Engineering & Technology, Dept of Biotechnology, Davanagere-577004; Bilachi S Ravindranath – Email: ravindranathbs@gmail.com; *Corresponding author

Received May 25, 2013; Revised June 03, 2013; Accepted June 07, 2013; Published July 12, 2013

Abstract:
Chlamydomphila pneumoniae, the causative agent of chronic obstructive pulmonary disease (COPD), is presently the fifth mortality causing chronic disease in the world. The understanding of disease and treatment options are limited represents a severe concern and a need for better therapeutics. With the advancements in the field of complete genome sequencing and computational approaches development have lead to metabolic pathway analysis and protein-protein interaction network which provides vital evidence to the protein function and has been appropriate to the fields such as systems biology and drug discovery. Protein interaction network analysis allows us to predict the most potential drug targets among large number of the non-homologous proteins involved in the unique metabolic pathway. A computational comparative metabolic pathway analysis of the host H. sapiens and the pathogen C pneumoniae AR39 has been carried out at three level analyses. Firstly, metabolic pathway analysis was performed to identify unique metabolic pathways and non-homologous proteins were identified. Secondly, essentiality of the proteins was checked, where these proteins contribute to the growth and survival of the organism. Finally these proteins were further subjected to predict protein interaction networks. Among the total 65 pathways in the C pneumoniae AR39 genome 10 were identified as the unique metabolic pathways which were not found in the human host, 32 enzymes were predicted as essential and these proteins were considered for protein interaction analysis, later using various criteria’s we have narrowed down to prioritize ribonucleotide-diphosphate reductase subunit beta as a potential drug target which facilitate for the successful entry into drug designing.

Key words: Chlamydomphila pneumoniae, COPD, computational approaches, metabolic pathway analysis, Protein interaction network analysis.

Background:
Chlamydomphila pneumoniae is an obligate intracellular Gram negative bacterium which belongs to member of Chlamydomphila. C pneumoniae infections are often persistent, and an acute infection may sometimes turn chronic [1]. Acute C pneumoniae infection is involved in the pathogenesis of different mammals and humans which cause bronchitis, emphysema and pneumonia [2]. It is been reported as TWAR organism as it was isolated from Taiwan in acute respiratory isolate, also known as AR-39. Serological studies indicate that almost many have developed antibodies against C pneumoniae; men exhibit more seroconversion than women. It is known for causing many of the chronic infection which is being actively investigated as a cause of several systemic diseases like COPD [3], Chronic asthma [4], Coronary artery disease [5], Atherosclerosis[6], Multisclerosis, Alzheimer’s disease [7] and Lung cancer[8] etc.
Similar to viruses, Chlamyphila pneumoniae is a parasitic organism that needs an in-vivo environment to reproduce and is therefore totally dependent on the host cell for survival. The C. pneumoniae life cycle provides ideal circumstances for the establishment of chronic infection. COPD have recently been associated with chronic C. pneumoniae infection [2], these chronic infections of respiratory tissues could contribute to the pathogenesis of COPD by altering the host response which has proven to be extremely difficult to diagnose and impossible to treat with current antibiotics. The prevalence of COPD has emerged as the major cause of morbity and mortality rate globally and it is anticipated as to become the third leading cause of death by 2020 and the 5th leading cause of loss of Disability Adjusted Life Year’s (DALYs) as per the Global burden of disease study (GBDS) [9], the region wise projection of developing countries like India were even worse.

The successful completion of the human genome project has revolutionized the field of drug-discovery to identify the biomarkers and to develop vaccines against human pathogens. Development of various computational tools to analyze the molecules from the genome to proteome level has made a major impact in the advancement of insilico based studies on which we can rely upon [10]. The basic strategy of identifying the non-homologous proteins which contribute to identify novel drug targets has been a preference to study host-pathogen interactions. The computational approach of studying newer molecular and genomics research techniques with the modern approach like comparative metabolic pathway analysis [11] and protein-protein interaction network studies is contributing to find precisely the important proteins/enzymes which interact and play a role of pathogenicity in many infectious and systemic diseases which can be concluded as potential drug targets [12]. Recently due to the development of novel computational tools, algorithms and methods, it’s been able to predict protein functions and protein interactions which are experimentally proved. The current study focuses on protein interaction network, since infections are usually initiated by protein-protein interactions. In our study we are able to predict single to multiple protein interaction events by using the available sophisticated tools.

**Methodology:**

*Analysis of host and pathogen metabolic pathways*

Whole genome-wide metabolic pathway analysis of host and pathogen was performed via the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [13]. Metabolic pathways and assigned identification numbers of the pathogen C. pneumoniae and the human host were extracted from the KEGG Pathway database. A manual sorting and comparison was then performed, and pathways that were not present in the host but were present in the pathogen, according to the KEGG database annotations, were identified as unique to C. pneumoniae, while the remaining pathways were considered as common pathways. C. pneumoniae enzymes from unique and common pathways were then identified, and the specific protein sequences were obtained from the NCBI GenPept database [14].

*Identification of Unique metabolic pathways and non-homologous proteins*

The Two-step analysis was performed between host and pathogen proteomes for identification of non-host proteins from C. pneumoniae. Firstly, manual sorting of pathways was performed then only proteins from pathogen-specific pathways were subjected to NCBI BLASTP analysis (http://blast.ncbi.nlm.nih.gov/Blast.cgi) [15] (Figure 1). Secondly, proteins from common pathways were also compared by BLASTP analysis. In each scenario, searching was restricted to proteins from human RefSeq protein database, by selecting the option provided under the NCBI BLASTP parameters. Proteins without hits below the cutoff E-value 0.005 and, 35% identity were selected as non-host pathogenic proteins.

**Identification of Non-host Essential Proteins**

In the present step all the non-host proteins were identified, they were subjected to study their essentiality by comparing the protein sequences of the C. pneumoniae against DEG database by using BLASTP search available specifically for prokaryotic organisms in DEG database at (http://tubic.tju.edu.cn/deg/) [16] for predicting putative non-host essential proteins using below E-value cutoff 0.005 and 35% identity.

---

**Figure 1:** Graph showing the prioritization and classification of proteins and unique pathways identified from the total metabolic pathways in Chlamyphila pneumoniae AR39.

**Graph Showing prioritization of drug targets**

| Total proteins in Chlamyphila pneumoniae | Identified proteins in Nonhomologous pathways | Identified proteins in protein interaction network | Prioritized drug targets |
|------------------------------------------|-----------------------------------------------|--------------------------------------------------|-------------------------|
| 119                                      | 170                                           | 32                                               | 32                      |
| 70                                       | 32                                            | 22                                               | 22                      |

**Graph showing prioritization and classification of proteins and unique pathways identified from the total metabolic pathways in Chlamyphila pneumoniae AR39.**

**Total metabolic pathways in Chlamyphila pneumoniae AR39.**

| Total metabolic pathways in Chlamyphila pneumoniae AR39 | Unique Pathways identified |
|--------------------------------------------------------|-----------------------------|
| 119                                                    | 10                          |

**Figure 1:** Graph showing the prioritization and classification of proteins and unique pathways identified from the total metabolic pathways in Chlamyphila pneumoniae AR39.
Protein-Protein interaction network prediction (PPI)

Protein-protein interactions play a pivotal role in many biological processes; prediction of these interactions gives an insight into identification of next gen therapeutic targets [12]. PPI peak analysis has been performed for identifying potential drug targets. In the present analysis we have been able to identify the most potential metabolic functional associations among all identified choke point proteins through protein interaction database STRING [17]. In STRING, methods namely neighborhood (green), co-occurrence (blue), gene fusion (red), co-expression (black) and experimental method (pink) have been used for potential metabolic interaction prediction.

Testing the druggability and Prioritization of Drug targets

Previous computational approaches have focused mainly on determining whether a protein is non-homologous to H.sapiens and its essentiality and in which pathway it is involved. Along with these important criteria’s, and with the advancement in genome sequencing, availability of computational tools, coupled with experimental data, indicates that there are several additional factors which are crucial in determining the feasibility of therapeutic targets. The most important criteria’s considered and predicted are: Molecular Weight (<110kDa) [18]; toxicity and localization [19]; absence and presence of transmembrane [20], availability of 3D structure [21] and druggability [22]. Assimilation of such additional information paves a way in prioritizing the potential therapeutic targets.

Results & Discussion:
Metabolic Pathway Analysis and Identification of Non-host Proteins

The present study focuses on in-silico based comparative metabolic pathway analysis of host H. sapiens and the pathogen C. pneumoniae. Firstly initial information about the metabolic pathways associated with C. pneumoniae was extensively analyzed, where KEGG pathway database was used as a source of information for metabolic pathways. H. sapiens and the pathogen C. pneumoniae were retrieved from the KEGG database. It presently contains information about 65 metabolic pathways in C. pneumoniae AR39 and 110 in H. sapiens. Protein names and total number of proteins present in each pathway were calculated, and comparative analysis was performed manually for the identification of pathways specific to C. pneumoniae. 10 different metabolic pathways were identified as unique to C. pneumoniae, and pathways were common to remaining C. pneumoniae strains. Secondly the corresponding protein sequences were retrieved from the NCBI database. BLASTP search was performed specifically against H. sapiens with the following criteria with the e-value threshold 0.005 and identity percentage of ≥35% [14] were considered as non-host proteins which do not have human homologues.

Finding Essential Genes

By taking the advantage of available essential genes information from DEG database (http://tubic.tju.edu.cn/deg/), we report 32 proteins from the C. pneumoniae genome as essential via DEG BLASTP search against prokaryotic organisms specifically, with E-value threshold cut-off of 10^-19, minimum bit score of 100, and identity percentage ≥35% [15]. Following our proteome-wide analysis, a total of 32 proteins were predicted as essential, out of the 170 total proteins in unique pathways where, 76 proteins were of single entry, and 94 proteins were of multiple entries, therefore they were predicted as essential in unique pathways.

Protein interaction network analysis

Functional associations between non- homologous proteins have been predicted by STRING database (http://string.embl.de/) [16]. Protein interaction network has been made between highly correlated (>0.9) proteins, where 32 protein interactions based on the above mentioned methods from which interactions has been predicted, and have been collectively considered for potential metabolic interaction analysis. The considered metabolically interacting proteins were found in DEG database of essential gene with their function and associated literature [16].

Figure 2: Showing nrdB (ribonucleotide-diphosphate reductase subunit beta) protein interacting with nrdA, adk, pyk, ndk, gmk, tmk, pnp, nrdR, CPn_0775 and dut proteins with evidence score ranging from 0.999-0.929 predicted by STRING.

Prioritization of Drug Targets

Based on above mentioned criteria’s, 21 proteins were bacterial exotoxins and 1 protein was an endotoxin. In our earlier work UDP-N-acetylglucosamine 1-carboxyvinyltransferase and Transketolase are reported [23] remaining targets are reported for the first time. We recommend a single potential drug target, ribonucleotide-diphosphate reductase subunit beta (Figure 2) after the final conclusion based on criteria’s met by the target, which is also a valid target of Cladribine an approved drug from Drugbank, also it is proposed in many other studies as cancer target [24, 25] and it plays a key role in DNA synthesis and cell growth control by reducing ribonucleotides to their corresponding deoxyribonucleotides which are crucial in DNA replication and repair, and is also found to be involved in pathogenicity as it is an exotoxin from our analysis.

Conclusion:

In the present study, comparative metabolic pathway approach and protein interaction network analysis of the causative agent of COPD has been performed to identify the unique pathways...
present in pathogen *C pneumoniae* which can be targeted for effective drug discovery and vaccine development. There are 10 unique pathways in Table 1 (see supplementary material) identified out of 65 pathways. Among them unique metabolic pathways consists 170 enzymes, in which 150 distinct non-host enzymes were identified. 32 different enzymes have been found as highly interacting metabolic proteins in *C pneumoniae* and their essentiality was confirmed by Database of essential genes which play pivotal role in bacterial pathogenicity and essential nutrient uptake. A list of 22 drug targets has been prioritized on the basis of number of pathogenic interactants. We specifically recommend ribonucleotide-diphosphate reductase subunit beta as the most prioritized potential drug target for drug designing and inhibition of *C pneumoniae* pathogenicity.

References:
[1] Matthias Krüll et al. Thromb Haemost. 2005 94: 319 [PMID: 16113821]
[2] Grayston JT et al. J Infect Dis. 1993 168: 1231 [PMID: 8228356]
[3] Blasi F et al. Clin Microbiol infect. 2009 15: 29 [PMID: 19220337]
[4] Johnston et al. Am j Respir crit care Med. 2005 172: 1078 [PMID: 15961690]
[5] Anderson et al. Circulation. 1999 99: 1540 [PMID: 10096928]
[6] Campbell et al. Nat Rev Microbiol. 2004 2: 23 [PMID: 15035006]
[7] Sriram et al. Ann neurol. 1999 46: 6 [PMID: 10401775]
[8] Littman et al. Cancer Epidemiolo Biomarkers Prev. 2005 14: 773 [PMID: 15824142]
[9] Shibuya et al. World Health. 2001
[10] Whittaker PA et al. Trends pharmacol sci. 2003 24: 434 [PMID: 12915054]
[11] Butt AM et al. PLoS One. 2012 7: e43080 [PMID: 22912793]
[12] Kushwaha SK et al. J Theor Biol. 2010 262: 284 [PMID: 19833135]
[13] Kanehisa M, Nucleic Acids Res. 2012 40: D109 [PMID: 22080510]
[14] http://www.ncbi.nlm.nih.gov/
[15] Altschul SF et al. Nucleic Acids Res. 1997 25: 3389 [PMID: 9254694]
[16] Zhang R et al. Nucleic Acids Res. 2009 37: D455 [PMID: 18974178]
[17] Franceschini A et al. Nucleic Acids Res. 2013 41: D808 [PMID: 23203871]
[18] http://web.expasy.org/compute_pi/
[19] http://www.imtech.res.in/raghava/btxpred/
[20] http://www.cbs.dtu.dk/services/TMHMM/
[21] www.rcsb.org/
[22] http://www.drugbank.ca/
[23] Bilachi S Ravindranath et al. Bioinformation. 2013 9: 506 (In press)
[24] Zhang M et al. Int J Gynecol Cancer. 2013 23: 659 [PMID: 23466567]
[25] Liu X et al. Clin Sci (Lond). 2013 124: 567 [PMID: 23113760]

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Edited by P Kangeane

Citation: Ravindranath et al. Bioinformation 9(12): 605-609 (2013)
**Supplementary material:**

| Sl No | Pathways Id | Unique Pathways                                |
|-------|-------------|------------------------------------------------|
| 1.    | cpa00540    | Lipopolysaccharide biosynthesis                 |
| 2.    | cpa00550    | Peptidoglycan biosynthesis                     |
| 3.    | cpa00624    | Polycyclic aromatic hydrocarbon degradation    |
| 4.    | cpa01110    | Biosynthesis of secondary metabolites          |
| 5.    | cpa01120    | Microbial metabolism in diverse environments   |
| 6.    | cpa02020    | Two-component system                           |
| 7.    | cpa02040    | Flagellar assembly                             |
| 8.    | cpa02060    | Phosphotransferase system (PTS)                |
| 9.    | cpa03070    | Bacterial secretion system                     |
| 10.   | cpa00680    | Methane metabolism                             |