

Functional Analysis of Hypothetical Proteins of *Chlamydia Trachomatis*: A Bioinformatics Based Approach for Prioritizing the Targets

Prashant K Mishra^{1,2}, Subash C Sonkar¹, Sree Rohit Raj¹, Uma Chaudhry² and Daman Saluja^{1*}

¹Dr. B. R. Ambedkar Center for Biomedical Research, University of Delhi, New Delhi, India

²Bhaskaracharya College of Applied Sciences, University of Delhi, New Delhi, India

Abstract

The various genome sequencing projects have led to the accumulation of entire set of gene sequences of many organisms. Among the sequenced genomes are numerous genes which code for proteins of unknown function. These genes are termed as hypothetical genes and their corresponding gene products are known as Hypothetical Proteins (HPs). Analyzing and annotating the functions of these HPs is important in pathogenic organisms such as *Chlamydia trachomatis* that causes various sequelae of diseases by infecting different sites in humans. Functional annotations of these HPs provides insights into their exact molecular function and may help in identification of novel drug or vaccine candidates for the control of infections caused by *C. trachomatis*. In the present study, entire set of 336 HPs of *C. trachomatis* were retrieved from NCBI and analyzed for their function using bioinformatics tools such as CDD-BLAST, PFAM, TIGRFAM and SCANPROSITE. The analysis revealed that some of the HPs possessed functionally important domains like protease, ligase, synthase, translocase and zinc finger domain. Some of the hypothetical proteins were found to be similar to transcriptional regulators while others were homologous to chaperonins. A few of the HPs corresponded to the bacterial secretory pathway proteins. The structural prediction of the annotated proteins has been performed which further substantiate the functional characterization results. Bioinformatics approach used in this study, including sophisticated sequence analysis, domain characterization and structural prediction studies, can provide a useful lead to experimentally annotate and corroborate these studies. Data generated by this study might facilitate swift identification of potential therapeutic targets and thereby enabling the search for new inhibitors or vaccines.

Keywords: *Chlamydia trachomatis*; Hypothetical proteins; Drug targets; Vaccine candidates

Introduction

Chlamydia trachomatis infection is one of the most common sexually transmitted diseases worldwide, with an estimated 7.2 million disease cases annually, of which majority of the cases occur in Asia (including India), Sub-Saharan Africa and South America [1-3]. Moreover, in India we have limited capacity to effectively screen for infections caused by *C. trachomatis* and only few labs are actively working in this area [4-9]. Since it is well established that chlamydial infections can lead to a variety of asymptomatic and symptomatic manifestations including vaginal muco-purulent discharge, endometritis, salpingitis and Pelvic Inflammatory Disease (PID) [10], it becomes imperative to screen for such infections at an early stage. In women, the infection is mostly asymptomatic and among the infected women, it is estimated that approximately 20% develop PID, 4% chronic pelvic pain, 3% become infertile and 2% have an adverse pregnancy outcome [11-13]. These complications increase further because of the occurrence of numerous *C. trachomatis* serovars, of which eight serovars (D, E, F, G, H, I, J, K) are known to cause genital infections [14].

In view of the observed complications, it is clear that not only screening programs are important, proper treatment regimens should also be followed for its effective control. In the pre-antibiotic era, chlamydial infections were often life threatening and even today they may give rise to death despite treatment with antibiotics as it has been established that some of the *C. trachomatis* serovars can produce a number of different components that may contribute to virulence, eventually leading to death of a patient [10]. Although, antibiotic therapy is thought to eliminate these chlamydial infections [15] and thereby prevent such extreme cases, it does not treat the established pathology. This, together with the fact that chlamydial infections

can often be asymptomatic, also point towards the urgent need for the development of preventative measures such as vaccination for control of the disease. Moreover, as per the guidelines of World Health Organization (WHO), treatment is largely based on symptomatic case management [1]. Since a large proportion of infective population remains asymptomatic, treatment on the basis of such guidelines not only misses out on asymptomatic patients but could also lead to over treatment. It is important to understand that treatment regimen should be based on local antibiotic sensitivities but such pattern is rarely known in resource limited settings such as India and hence may contribute to emergence of resistant strains of this pathogen which may further spread to other countries as well. Even though, today most of the antibiotics are effective against *C. trachomatis* infections but the time is not far wherein the antibiotics would turn out to be ineffective as with other infectious diseases. Therefore, it is becoming increasingly important that we not only explore novel drugs / targets for its treatment but also develop new vaccine candidates as a step towards prevention of the disease. Unfortunately, development of novel anti-chlamydial drugs is hampered because of our lack of understanding of its pathogenesis. Moreover, *C. trachomatis* infection can be prevented

***Corresponding author:** Daman Saluja, Dr. B. R. Ambedkar Center for Biomedical Research, University of Delhi, New Delhi, India, E-mail: Delhidsalujach59@gmail.com

Received November 12, 2013; **Accepted** November 27, 2013; **Published** December 07, 2013

Citation: Mishra PK, Sonkar SC, Raj SR, Chaudhry U, Saluja D (2013) Functional Analysis of Hypothetical Proteins of *Chlamydia Trachomatis*: A Bioinformatics Based Approach for Prioritizing the Targets. J Comput Sci Syst Biol 7: 010-014. doi:10.4172/jcsb.1000132

Copyright: © 2013 Mishra PK, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

worldwide, once we are able to develop newer vaccines as current vaccines are no longer effective against Chlamydial infections owing to the presence of multiple serovars. These two objectives, i. e., novel drug targets and / or vaccine development could be readily achieved if one is able to understand Chlamydial biology in a holistic manner. Along with experimental studies, bioinformatics and genomics might play an important role in achieving the above mentioned objectives, as they have done for other infectious organisms. Over the last decade, more than 150 complete genomes of diverse bacteria, archaea and eukaryotes have been sequenced and many more are currently in pipeline [16,17]. Although, a large number of microbes have been sequenced, about 50% of genes do not show homology to functionally characterized proteins. Functional annotation of these HPs might offer an opportunity to understand novel drug designing methods and vaccine development.

In the present study, we have made an attempt to exploit the data generated using bioinformatics tools to help us investigate newer treatment options / preventive measures available for the pathogenic organism, *C. trachomatis*. The completed genome sequence for *C. trachomatis* D/UW-3/CX provided us with a comprehensive inventory of all the proteins potentially produced by this organism. A significant question that we wanted to answer was as to how this information can guide us in identifying novel drug targets or vaccine candidates. The Chlamydial whole genome is a single chromosome which contains 1042519 base pairs (bps) [~1.04 megabase pairs (MB)] having Adenine and Thymine (AT) content of 58.7%. Total number of protein encoding genes throughout the *C. trachomatis* genome is 894 (genome sequencing and annotation methodologies available are at the Science website: (www.sciencemag.org/feature/data/982604/.shl)). Out of these 894 protein coding genes, 558 (62%) genes encode for proteins that have assigned function and 336 (38%) genes are hypothetical in nature.

Hypothetical Protein (HP) is a protein that is predicted to be expressed from an open reading frame, but for which there is no experimental evidence of translation. These constitute a substantial fraction of proteomes of prokaryotes and are even present in eukaryotes [16]. With the general belief that the majority of HPs are the product of pseudogenes, it is essential to have a tool with the ability of pinpointing the minority of HPs with a high probability of being expressed. In the strict sense, HPs are predicted proteins, as predicted from their nucleic acid sequences and that have not been shown to be expressed by experimental procedures. Moreover, these proteins are characterized by low identity to known, annotated proteins. Among these HPs, a separate class of “conserved hypothetical” proteins are defined as a large fraction of genes in sequenced genomes encoding those that are found in organisms from several phylogenetic lineages but have not been functionally characterized and described at the protein chemical level [18]. These structures may represent up to half of the potential protein coding regions of a genome.

In order to treat infectious diseases such as those caused by *C. trachomatis*, functional annotation of these HPs might open avenues for prioritizing vaccine candidate genes or novel drug targets. Structural genomics initiatives provide ample structures of hypothetical proteins at an ever increasing rate. However without functional annotation of these proteins, this structural analysis would be of no use to biologists who are always interested in deciphering particular molecular mechanisms. Moreover, some of the proteins, which are considered to be well annotated, may have additional functions beyond their listed records. Undertaking such studies, a series of additional protein pathways and cascades can be revealed, completing our fragmentary

knowledge on the mosaic of proteins *per se*. Lastly we may emphasize that the analysis of HPs would be of benefit to genomics enabling the discovery of so far unknown or even predicted genes. These annotated HPs may serve as markers and pharmacological targets in the era of personalized medicine for *C. trachomatis*.

Methods

Sequence retrieval and functional annotation

Complete genome sequence of *C. trachomatis* D/UW-3/CX was retrieved from NCBI database (Sequence and annotation available at (<http://chlamydia-www.berkeley.edu>) and GenBank under accession number AE001273) and the sequences of “hypothetical proteins” of *Chlamydia trachomatis* were analyzed. Out of 894 protein sequences, 336 hypothetical protein sequences were analyzed for the presence of conserved domains using sequence similarity search with orthologous family members available in various databases using web-tools.

Four bioinformatics tools CDD-BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) [19-21], TIGRFAM (<http://www.jcvi.org/cgi-bin/tigrfams/index.cgi>) [22], PFAM (<http://www.pfam.sanger.ac.uk/>) [23] and SCANPROSITE (<http://prosite.expasy.org/scanprosite/>) [24] were used, which can search the defined conserved domains in the targeted protein sequences and further assist in the classification of putative proteins in a particular protein family.

CDD-BLAST is NCBI's web interface to search the Conserved Domain Database with protein query sequences. It uses RPSBLAST, a variant of PSI-BLAST, to quickly scan a set of pre calculated Position-Specific Scoring Matrices (PSSMs) with a protein query. PFAM is a collection of multiple protein-sequence alignments and Hidden Markov Models (HMMs) and provides a good repository of models for identifying protein families, domains and repeats. TIGRFAM is manually curated database which exploits HMMs to annotate protein families by search alignment for each family. It contains more than 1600 protein families and gives a base through which one can extract a number of information regarding genome annotation, sequence similarity, sequence alignment, function profiling and phylogenetics profile analysis during model construction. SCANPROSITE is an improved version of web based tool which detect PROSITE signature matches in given protein sequence.

Functional categorization

Hypothetical proteins analyzed by the above mentioned function prediction web tools have shown the variable results when searched for the conserved domains in hypothetical sequences and different confidence levels have been generated on the basis of collective results of these web-tools.

1. If the given four tools indicate the same functions then the confidence level were to be 100 percent.
2. If the given three tools indicate the same functions and one is showing different function then the confidence level were to be 75 percent.
3. If the given two tools indicate the same functions and two are showing different functions then the confidence level were to be 50 percent.
4. If only one tool indicates the function and other tools are showing different functions then the confidence level were to be 25 percent.

Protein structure prediction

Predicting the 3D structure of a HP from its amino acid sequence using computational methods also allow us to predict its function to an extent [25]. Two types of tertiary structure prediction methods are known namely, *ab initio* methods which predict a protein structure based on physico-chemical principles directly, and template-based methods, which uses known protein structures as templates. Template based methods include homology or comparative modeling, and fold recognition via threading.

In our study we have used an online server PS2-v2 (PS Square version 2) Protein Structure Prediction Server [26] which is a template based method to predict the structure of HPs. PS2-v2 uses the strategies of Pair-wise and multiple alignments by combining powers of the programs PSI-BLAST, IMPALA and T-COFFEE in both target – template selection and target–template alignment and finally it constructs the protein 3D structures using integrated modeling package of PS2 using best scored orthologous template.

We found that (PS)2-v2 server to be a very user-friendly web server. The query protein sequence of the HP was given as an input in FASTA format. The server provided three modes (Automatic, Manual and ‘Use this template’) for choosing the template. The default mode was ‘Automatic’. In the automatic mode, (PS)2-v2 automatically selects the modeling template(s). For the ‘Manual’ mode, server enables users to assign specific template(s) from a list of candidates. The ‘Use this template’ mode allows users to assign a specific protein structure as the template. We used the default mode for all the eleven proteins in order to predict their structures. Predicted results were received by us later at our email address. Output of the server showed a list of templates, selected template(s), target-template alignment(s), predicted structure(s) and structure evaluations as the result.

Results and Discussion

Advances in high throughput modern DNA sequencing technologies combined with its cost efficiency has enabled sequencing of a large number of bacterial genomes. Since many of the genes are conserved across a large number of bacterial genomes, accurate annotation of the genes usually relies on sequence homology methods wherein, the function of a specific gene is assigned based on sequence similarity to gene with known function [27]. In spite of this enormous proliferation of genomic data more than one third of genes have no assigned function. One of the reasons is the functional divergence of similar sequences during the course of evolution [28,29]. Hence sequence homology based methods alone fail to assign accurate functions to a large number of genes and may lead to imprecise annotations [30]. To circumvent this problem, multiple tools should be used to assign functions to hypothetical proteins, as this may help in reducing the fraction of HPs considerably. The present study concentrated on the functional annotation of hypothetical proteins from *C. trachomatis* using diverse bioinformatics tools. The four web-tools used in the current study helped us to search the presence of conserved domains in 336 Hypothetical Proteins (HP). Domains are often identified as recurring (sequence or structure) units, which may exist in various contexts and can be thought of as distinct functional and/or structural units of a protein. In molecular evolution such domains may have been utilized as building blocks and may recombine in different arrangements to modulate protein function [31]. Once the presence of a specific domain in a protein was established, we further classified proteins into various categories. Depending on our results,

as represented in Table S1, we found that there were 11 proteins having a consensus in the presence of the domain using all the four Bioinformatics tools and hence they were grouped together under the confidence limit of 100%, as shown in Table 1.

Out of the remaining 325 proteins, we could not find any specific domains for 83 proteins using the four bioinformatics tools. For these 83 proteins, structural analysis might provide some meaningful results. For other hypothetical proteins (n=242), specific domains were identified using one, two or three of the above mentioned tools. Accordingly these were categorized under the confidence limit of 25% for 134 proteins, 50% for 66 proteins and 75% for 42 proteins (Table 1). These may or may not be specific for the highlighted domains. The exact function of these proteins requires further studies.

Within a protein family, a domain or fold may be more strongly conserved than the entire sequence [32]. Amongst the 11 HPs for which functional domain was identified with a confidence level of 100%, three proteins (CT110, CT341, CT396) were classified as chaperonins (Table 2). Other proteins showed domains suggestive of their function as ligase, synthase, protease, bacterial secretory pathway proteins, translocase and a zinc finger domain carrying protein suggesting that these proteins of *Chlamydia* may be involved in performing similar function. Presence of zinc finger domain in CT407 implies that it might be a DNA binding protein, probably a transcription regulator. The classification of the HPs according to the presence of specific domains and their super-family descriptions is given in Table 2.

Biochemical function of the protein can also be verified by inclusion of structural information [33]. The three dimensional structures of the eleven annotated proteins were modeled by PS square (PS2-v2) online server, which is dependent on Template-Based Modeling (TBM) and fold recognition methods. Using this method, a prediction model is built based on the coordinates of the appropriate template of the protein. These approaches generally involve four steps: 1) a representative protein structure database is searched to identify a template that is structurally similar to the protein target; 2) an alignment between the target and the template is generated that should align equivalent residues together as in the case of a structural alignment; 3) a prediction structure of the target is built based on the alignment and the selected template structure, and 4) model quality evaluation. The first two steps significantly affect the quality of the final model prediction in TBM methods. The templates used by the server to model these proteins are tabulated in Table 3. The modeling of CT110, CT341, CT396 proteins using PS square (PS2-v2) online server further substantiated the chaperonin function of these three proteins

Percentage of similarity	0%	25%	50%	75%	100%
Number of hypothetical proteins (Total 336)	83	134	66	42	11

Table 1: Confidence level of All Four Tools Used in this Study (Total proteins 336).

S. No.	Functional categories to which they belong	List of proteins
1.	Chaperonin	CT110, CT341, CT396
2.	Ligase	CT146
3.	Synthase	CT171, CT257
4.	Protease	CT431, CT706
5.	Bacterial secretory pathway proteins	CT571
6.	Translocase	CT701
7.	Zinc finger protein	CT407

Table 2: Putative Functional Category of Hypothetical Proteins with 100% confidence level According to the Functional Domain They Contain (Total 11 Proteins).

(Tables 2 and 3). Of these eleven proteins, only one protein showed discrepant results. CT257 modeled as a transporter protein, whereas it was predicted to be a synthase using web-based functional prediction tools. Further experimental studies or genome context based methods might prove helpful in identifying the exact function for this protein. These structure prediction results further substantiate the annotated function of the remaining ten proteins.

Based on these results, we emphasize that publically available bioinformatic tools can provide a general function of a few of the HPs which can be further used as a lead for designing experimental approaches geared towards evaluation of exact function of the gene. Although at present the exact function of many genes is still inscrutable, detailed sequence analysis along with genome context methods (such as operon analysis) and expression analysis may provide useful clues as to their cellular role. Genomic context methods have been recently explored and are designed to detect presumed functional constraints on genome evolution. They predict functional associations between protein coding genes by analyzing gene fusion events, the conservation of gene neighborhood, or the significant co-occurrence of genes across different species [34-38]. Unlike homology based annotation, which infers molecular features by information transfer from experimentally characterized proteins, genomic context methods predict functional associations. Therefore, it should be noted that these methods do not provide information about exact biochemical or experimental function and are time consuming. *In silico* analysis described here provide an easy and accurate method of assigning function for various HPs. Nevertheless using systematic hierarchical approaches together, would help scientists in reducing the number of uncharacterized HPs.

Although we have been able to accomplish a comprehensive *in silico* analysis, still function of around 300 genes of the studied pathogenic bacteria *C. trachomatis*, which produce many virulence factors and cause serious infections and disease complications, is still ambiguous. Further understanding of functional properties of these HPs of *C. trachomatis* will not only provide a better insight into the pathogenesis of *C. trachomatis* but may also help in identifying novel therapeutic candidates. In fact, there has been a recent report wherein a novel genomics approach using Codon Adaptation Index (CAI), a measure that is used to predict the translational efficiency of a gene based on synonymous codon usage, is coupled with subtractive genomics approach for mining potential drug targets [39]. Using the strategy, the group was able to identify 8 potent target genes from *Streptococcus pneumonia* and *Haemophilus influenzae*, which were found to be functionally significant. In comparison to their approach,

our study facilitates swift identification of the hidden function of HPs which could become potential therapeutic targets and may indicate a key role in host-pathogen interactions. Once established as novel drug or vaccine targets, further research for new inhibitors and vaccines can be accomplished.

Conclusions

In the present study, we have taken a close look at 336 chlamydial HPs whose function was analyzed using diverse publically available bioinformatics tools. Based on the various domains studies, we were able to classify 11 hypothetical proteins under different functional categories. The analysis revealed functionally important domains and families which were involved in inducing protein synthesis and multiple antibiotic resistances in the bacteria and also perform enzymatic functions. A few of the annotated proteins could turn out to be novel targets to combat antibiotic resistance and vaccine development once their role is validated experimentally. Future studies may be directed towards subcellular localization of these annotated proteins for elucidating their cellular processes. Knowledge of the subcellular localization of a protein can significantly improve target identification during the drug discovery process. Once the structures of these annotated proteins is established and their function is known, further investigation into their ligand binding sites would help us identify newer antimicrobials against resistant strains. Currently there is also an immediate need for newer and effective vaccine providing protection against *Chlamydia trachomatis* due its inherent antigenic variation. Proteins that are secreted by various microbes to the extracellular environment could turn out to be useful antigens that might induce protective immunity or could elicit an immune response of diagnostic value. Above mentioned subcellular localization studies would be, therefore, beneficial in identifying such antigenic proteins. Finally we may emphasize that quantitative computational analysis, carried out in the present study, may help us in better understanding of the biology of *Chlamydia trachomatis* as a whole and identify potential therapeutic leads at the molecular level.

Acknowledgments

Fellowship (UGC-NET) from University grant Commission to P.K.M and S.C.S.is acknowledged. Work was carried out at DBT-BIF facility at Dr. B. R. Ambedkar Center for Biomedical Research, University of Delhi.

References

- Global incidence and prevalence of selected curable sexually transmitted infections (2008) World Health Organization.
- Sexually Transmitted Infections (STIs) (2013) World Health Organization.
- Behets FM, Miller WC, Cohen MS (2001) Syndromic treatment of gonococcal and chlamydial infections in women seeking primary care for the genital discharge syndrome: decision-making. Bull World Health Organ 79: 1070-1075.
- Gopalkrishna V, Aggarwal N, Malhotra VL, Koranne RV, Mohan VP, et al. (2000) Chlamydia trachomatis and human papillomavirus infection in Indian women with sexually transmitted diseases and cervical precancerous and cancerous lesions. Clin Microbiol Infect 6: 88-93.
- Singh V, Salhan S, Das BC, Mittal A (2003) Predominance of Chlamydia trachomatis serovars associated with urogenital infections in females in New Delhi, India. J Clin Microbiol 41: 2700-2702.
- Sachdeva P, Patel AL, Sachdev D, Ali M, Mittal A, et al. (2009) Comparison of an in-house PCR assay, direct fluorescence assay and the Roche AMPLICOR Chlamydia trachomatis kit for detection of *C. trachomatis*. J Med Microbiol 58: 867-873.
- Patel AL, Sachdev D, Nagpal P, Chaudhry U, Sonkar SC, et al. (2010) Prevalence of Chlamydia infection among women visiting a gynaecology outpatient department: evaluation of an in-house PCR assay for detection of Chlamydia trachomatis. Ann Clin Microbiol Antimicrob 9: 24.

S. No.	Proteins	(PS)2-v2, 2009	
		Domain ID	Structure
1.	CT110	1we3F	Chaperonin
2.	CT146	2owoA	Ligase
3.	CT171	2ekcB	Synthase
4.	CT257	2yvyA	Transporter
5.	CT341	1nltA	Chaperonin
6.	CT396	2v7yA	Chaperonin
7.	CT407	1tjlA	Zinc Finger Protein
8.	CT431	1yg6A	Protease
9.	CT571	1p9rA	Secretory protein
10.	CT701	2vdaA	Translocase
11.	CT706	1yg6A	Protease

Table 3: Structural prediction of annotated proteins using the (PS)2-v2: template-based protein structure prediction server.

8. Jha HC, Srivastava P, Vardhan H, Singh LC, Bhengraj AR, et al. (2011) Chlamydia pneumoniae heat shock protein 60 is associated with apoptotic signaling pathway in human atheromatous plaques of coronary artery disease patients. J Cardiol 58: 216-225.
9. Bhatla N, Puri K, Joseph E, Kriplani A, Iyer VK, et al. (2013) Association of Chlamydia trachomatis infection with human papillomavirus (HPV) & cervical intraepithelial neoplasia - a pilot study. Indian J Med Res 137: 533-539.
10. Stephens RS (2003) The cellular paradigm of chlamydial pathogenesis. Trends Microbiol 11: 44-51.
11. Paavonen J, Eggert-Kruse W (1999) Chlamydia trachomatis: impact on human reproduction. Hum Reprod Update 5: 433-447.
12. Mårdh PA (2004) Tubal factor infertility, with special regard to chlamydial salpingitis. Curr Opin Infect Dis 17: 49-52.
13. van Valkengoed IG, Morré SA, van den Brule AJ, Meijer CJ, Bouter LM, et al. (2004) Overestimation of complication rates in evaluations of Chlamydia trachomatis screening programmes--implications for cost-effectiveness analyses. Int J Epidemiol 33: 416-425.
14. Igietseme JU, Ward ME (2004) Chlamydia update. Expert Rev Vaccines 3: 639-642.
15. Ridgway GL (1997) Treatment of chlamydial genital infection. J Antimicrob Chemother 40: 311-314.
16. Koonin EV, Galperin MY (2003) Sequence-Evolution-Function. Computational Approaches in Comparative Genomics. Kluwer Academic Publishers.
17. Bernal A, Ear U, Kyripides N (2001) Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. Nucleic Acids Res 29: 126-127.
18. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, et al. (1998) Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science 282: 754-759.
19. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.
20. Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, et al. (2007) CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Res 35: D237-D240.
21. Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755-763.
22. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, et al. (2010) The TIGRFAMs database of protein families. Nucleic Acids Research 38: D211-D222.
23. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, et al. (2002) The Pfam families database. Nucleic Acids Res 30: 276-280.
24. de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, et al. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res 34: W362-365.
25. Dong Xu, Ying Xu and Uberbacher EC. (2000) Computational tools for protein modeling. Current Protein and Peptide Science. 1:1-21.
26. Chen CC, Hwang JK, Yang JM (2009) (PS)2-v2: template-based protein structure prediction server. BMC Bioinformatics 10: 366.
27. Lee D, Redfern O, Orengo C (2007) Predicting protein function from sequence and structure. Nat Rev Mol Cell Biol 8: 995-1005.
28. Gerlt JA, Babbitt PC (2000) Can sequence determine function? Genome Biol 1: Reviews 0005.
29. Saghatelian A, Cravatt BF (2005) Assignment of protein function in the postgenomic era. Nat Chem Biol 1: 130-142.
30. Schnoes AM, Brown SD, Dodevski I, Babbitt PC (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. PLoS Comput Biol 5: e1000605.
31. Galperin MY and Koonin EV (2004) Conserved hypothetical' proteins: prioritization of targets for experimental study. Nucleic Acids Res. 32: 5452-5463.
32. Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. EMBO J. 5: 823-826.
33. Skolnick J, Fetrow JS (2000) From genes to protein structure and function: novel applications of computational approaches in the genomic era. Trends Biotechnol 18: 34-39.
34. Dandekar T, Snel B, Huynen M, Bork P (1998) Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem Sci 23: 324-328.
35. Enright AJ, Iliopoulos I, Kyripides NC, Ouzounis CA (1999) Protein interaction maps for complete genomes based on gene fusion events. Nature, 402: 86-90.
36. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, et al. (1999) Detecting protein function and protein-protein interactions from genome sequences. Science 285: 751-753.
37. 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 U S A 96: 2896-2901.
38. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proc Natl Acad Sci U S A 96: 4285-4288.
39. Vetrivel U, Subramanian G, Dorairaj S (2011) A novel *in silico* approach to identify potential therapeutic targets in human bacterial pathogens. Hugo J 5: 25-34.