WORLD JOURNAL OF PHARMACY AND PHARMACEUTICAL SCIENCES
SJIF Impact Factor 6.647
Volume 6, Issue 9, 70-80 Research Article ISSN 2278 – 4357

A COMPUTATIONAL METHOD FOR IDENTIFICATION OF PATHOGENIC CANDIDATES AMONGST THE HYPOTHETICAL PROTEINS IN CHLAMYDIA TRACHOMATIS D/UW-3/CX

Shahid Ali[a*], Laishram Singh[b], Reetika Arora[c], Yonghua Wang[a]

[a]Department of Food Science and Engineering, South China University of Technology, Guangzhou, China.

[b]Dr. B.R Ambedkar Center for Biomedical Research, University of Delhi, 110007, India.

[c]Amity Institute of Biotechnology, Amity University, Noida, India.

Article Received on
01 July 2017,

Revised on 22 July 2017,
Accepted on 12 August 2017,

DOI: 10.20959/wjpps20179-10003

***Corresponding Author**
Shahid Ali
Department of Food Science and Engineering, South China University of Technology, Guangzhou, China.

ABSTRACT

Chlamydia trachomatis is the most usual manifestation of curable bacterial sexually transmitted infection worldwide. It is primarily present as urethritis in males and end-cervicitis in females. Traditionally, tissue culture was the gold standard for diagnosis. However, with the advent of advanced molecular methods of diagnosis, which are not only highly sensitive and specific but are cost-effective too. However, due to generous size of genome analysis becomes an arduous task. This problem can be overcome by using bioinformatics tools which makes it possible to analyze large molecular data within a short period of time and cost effective. The size of C. trachomatis genome (AC: NC_000117.1) is relatively smaller in comparison to other bacteria and contains 887 functional proteins, out of which 266 are classified as hypothetical proteins (HPs) because of the unavailability of experimentally validated functions. The function of the HPs were predicted by integrating a variety of protein classification systems, motif discovery tools as well as methods that are based on characteristic features obtained from the protein sequence. The probable virulence factor proteins of HPs were predicted successfully. Furthermore, the virulent HPs present in the set of 18 functionally annotated proteins were predicted by using the Bioinformatics tools and the conformational behaviours of the proteins with highest virulence scores amongst the effector-1 protein were studied by using the molecular dynamics simulations. This study will

facilitate in a better understanding of various drug resistance and pathogenesis mechanisms present in the C. trachomatis which can be utilized in designing improved therapeutic agents.

KEY WORDS: Hypothetical proteins (HPs).

1. INTRODUCTION

Chlamydia trachomatis is one of the most common sexually transmitted bacterial infections worldwide, and women carry the brunt of this disease. Affected women are a potential source of infection to their partners. It causes urethritis in men and mucopurulent cervicitis, urethritis, and endometritis in women. Mucopurulent cervicitis can lead to at least three types of complications- ascending intraluminal spread of organism from cervix producing pelvic inflammatory disease (PID), ascending infection during pregnancy resulting in premature rupture of the membrane, chorioamnionitis, premature delivery and puerperal and neonatal infections (conjunctivitis and possibly interstitial pneumonia); and also an intensified risk of the development of cervical carcinoma[1]. A 3- to 4-fold increased risk of transmission of HIV is a continued cause of concern. The incidence of chlamydial infections has gone up substantially from 79 to 467 per 100,000 women between 1987 and 2004. According to the World Health Organization (WHO), 101 million chlamydial infections are detected annually worldwide. The clinical presentation, course, complications and late sequela of C. trachomatis closely resemble Neisseria gonorrhoeae infection.

C. trachomatis is also considered to be a leading cause of PID and female infertility worldwide. More than 13.5 per cent women under the age of 25 years, infected with C. trachomatis have a lower incidence of genital tract infection, a reduction up to 4.4 per cent in women 25 year and above[2]. In USA, approximately 20-30 per cent of PID cases have been attributed to C. trachomatis [3]. Recent studies from India have revealed the prevalence of C. trachomatis infection to be 23 percent in gynecology outpatient department (OPD)[4] and 19.9 per cent in STD patients[5] . Information retrieved from 30-60 per cent cases of salpingitis and PID[5] patients in India, while sero-prevalance is shown to be higher in at least one recent study[6]. An estimated 15-40 per cent of women with cervical chlamydial infections develop PID[7]. Twenty per cent of women who develop PID become infertile, 18 per cent develop chronic pelvic pain, and nine per cent have a tubal pregnancy[8]. There is a high frequency of the infection getting passed on to the fetus through the contaminated birth canal during parturition. Screening young women for signs of Chlamydia has been proven to be a cost-effective method of preventing PID. The US Preventive Services Task Force (USPSTF)

recommends that all women aged 24 years and above receive routine screening for Chlamydia. However, there is no sufficient evidence that suggests routine screening for chlamydial infection in asymptomatic males[9]. The challenge posed in the control of chlamydial disease is: as many as 70-80 per cent of women and up to 50 per cent men has asymptomatic infection[10]. This results in a large reservoir of unrecognized, infected individuals who are capable of transmitting the infection to their sexual partners. Further, the sequelae of C. trachomatis infection in women, namely PID, infertility and ectopic pregnancy, is the disastrous outcome of any STD except HIV or AIDS[11]. This paper will focus on annotation of hypothetical protein and how these Hp's are involved in STD disease?

2. MATERIAL AND METHODS

The computational framework for the functional annotations of Bacterial HPs was established by using the knowledge obtained from previous in-silico studies[12]. The elementary steps of the adopted pipeline are illustrated in Fig.1

2.1 Potential Target Selection

The genome of C. trachomatis contains 887 functional proteins and by utilizing the keyword search in the NCBI database; around 266 proteins were classified as HPs (http: //www.ncbi.nlm. nih.gov/genome/) and all the hypothetical proteins were extracted from the protein table. The Uniprot IDs and amino acid sequences of each HP were retrieved by using the available ID mapping tools (http://www.uniprot.org/). This tool helps in ID mapping of HPs proteins whose sequences we want to retrieve from the Uniport Database.

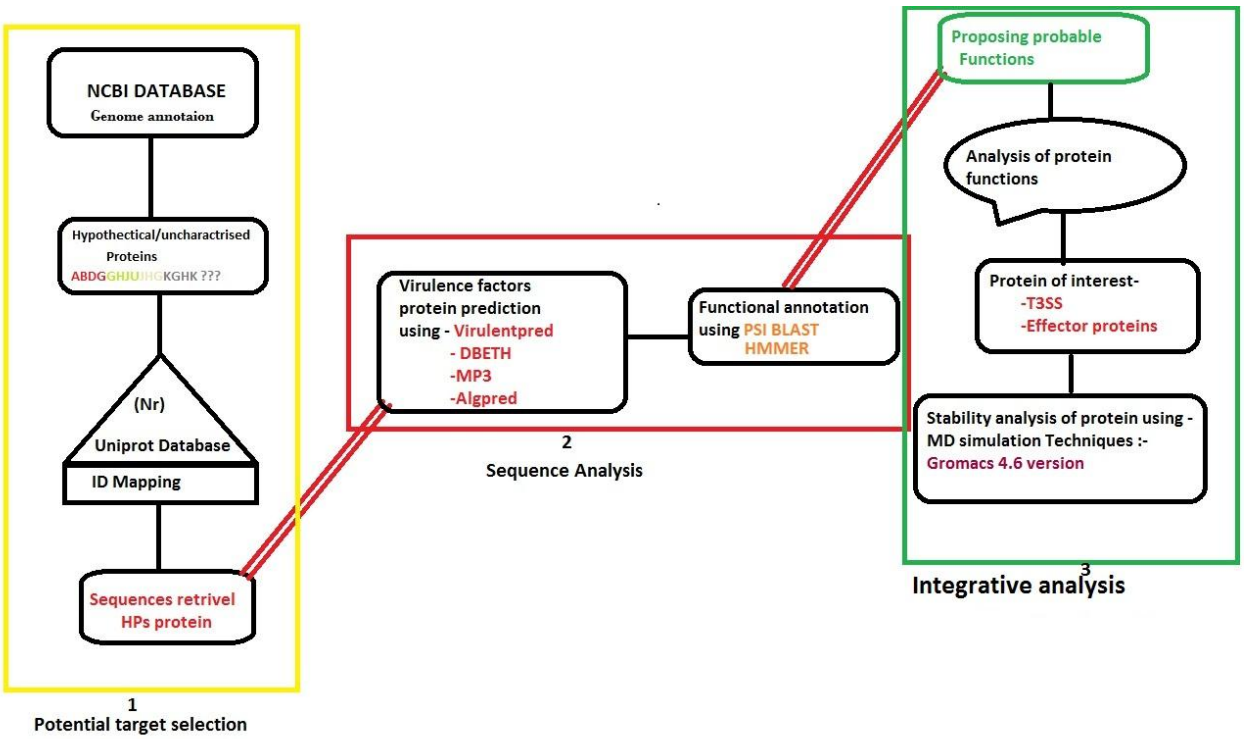

Figure 1. Pipeline for the identification of pathogenic candidates amongst the hypothetical proteins. Pipeline divides in three phases 1. Potential target selection 2. Sequence analysis 3. Integrative analysis.

2.2 Sequence analysis

In our study, first, we removed those HPs which were not involved in virulence factors amongst the 266 hypothetical proteins. The virulence factors (VFs) are essential for severity of infection caused by the pathogen[13], therefore, they are considered as potential targets in the drug discovery. A variety of uncharacterized or hypothetical proteins VFs were classified by using 4 online servers: VirulentPred server, DBETH (Database of Bacterial Exotoxin for Human), MP3 (Prediction of Pathogenic), Algpred (Predictions of Allergic Proteins and Mapping of IgE Epitopes)

2.3 Functional Annotation of HPs

The primary step for functional annotations of proteins involved sequence similarity searches in diverse biological databases. The sequence similarity search methods such as BLAST[14] , and HMMER[15] were used to identify functional homologues in the databases. In the PSI-BLAST search, the HPs with low sequence identities (<30%) as well as low query coverage (<50%) were excluded. The proteins that showed low sequence identities (<26%) were considered as remote homologues, while those with high sequence identities (>40%) were

referred to as close homologues. In a search, if the desirable magnitudes of the respective parameters were obtained, then it is considered as the possible hit for a particular HP. We assumed the probable function to an HP, if the respective functionality appeared maximum number of times with an e-value (<0.005). This procedure was adopted for every HP in order to obtain reliable results in similarity search (table 2). Similarly, HMMER was used for searching sequence databases for sequence homologues, and for sequence alignments. It implements methods using probabilistic models called profile hidden Markov models (profile HMMs). HMMER is often used together with a profile database, such as Pfam or many of the databases that participates in Interpro.

2.4 Molecular Modeling

The HPs with the highest predicted virulence scores amongst the effector_1 protein was selected for further structural analysis. The templates for the selected HPs were identified by using BLAST module of Discovery Studio [16]. In case of low sequence homology (<30%), the 3D structures were predicted by using the Ab-initio algorithms of the I-TASSER[15]. The evaluations of the predicted models were performed by using TM scores[17] and RMSD (root mean square deviations) values, calculated by the I-TASSER.

2.5 Model Stability Study

The best evaluated models were selected for energy minimization and stability studies. The modeled structures of virulent HPs were simulated by using GROMACS package (version 4.6.5)[18] in explicit solvent conditions at 300 K. In order to improve the electrostatic interactions, the virulent HPs were solvated in the Single Point Charge (SPC216) water model and simulated by using the Particle–Mesh–Ewald (PME) summation under Periodic Boundary Conditions (PBC). An initial structure of the protein was energetically minimized with a convergence criterion of 0.005 kcal/mol within 5000 steps of steep descent algorithms. The minimized structures were equilibrated for the time scale of 100 ps by using NVT and NPT ensemble conditions. The MD simulations were performed for 10 ns time scale by using the LINCS algorithm of GROMACS with a time step of 2 fs. All the resulting trajectories of the MD simulation were analyzed using the utilities present in the GROMACS package and Grace Plotter.

3. RESULT AND DISCUSSION

By using the available Bioinformatics tools, the sequences of 266 HPs were extensively analyzed (supplementary data) and 18 HPs were successfully predicted as virulence factors that are listed in Table 1.

Table (1): List of Virulence factors protein amongst the 266 HPs of Chlamydia trachomatis.

S.NO	Uniprot ID	CASCADE	Virulent pred	Algpred	DBETH	MP3 prediction
1	O84052	1.0283	Virulent	Allergen	Toxin	Pathogenic
2	O84053	1.0058	Virulent	Allergen	Toxin	Pathogenic
3	O84054	0.7794	Virulent	Allergen	Toxin	Pathogenic
4	O84144	1.0729	Virulent	Allergen	Toxin	Pathogenic
5	O84145	1.019	Virulent	Allergen	Toxin	Pathogenic
6	O84146	1.0257	Virulent	Allergen	Toxin	Pathogenic
7	O84251	0.8852	Virulent	Allergen	Toxin	Pathogenic
8	O84432	0.9897	Virulent	Allergen	Toxin	Pathogenic
9	O84482	0.9646	Virulent	Allergen	Toxin	Pathogenic
10	O84624	0.9929	Virulent	Allergen	Toxin	Pathogenic
11	O84625	1.015	Virulent	Allergen	Toxin	Pathogenic
12	O84626	0.9389	Virulent	Allergen	Toxin	Pathogenic
13	O84627	1.0957	Virulent	Allergen	Toxin	Pathogenic
14	O84659	1.1483	Virulent	Allergen	Toxin	Pathogenic
15	O84671	1.0572	Virulent	Allergen	Toxin	Pathogenic
16	O84708	1.053	Virulent	Allergen	Toxin	Pathogenic
17	O84855	1.0681	Virulent	Allergen	Toxin	Pathogenic
18	O84856	1.0703	Virulent	Allergen	Toxin	Pathogenic

The functions of 11HPs were successfully predicted that are listed in (Tables 2). Furthermore, the p-value for functionality assignment was set to 0.05, as the probability of random function predictions was assumed to be 0.05. The functional analyses through the adopted pipeline enabled us to comprehend the advantages and disadvantages of the available methods.

Table (2): List of functionally annotated HPs of Chlamydia trachomatis.

S.NO	Uniprot ID	Functions
1	O84052	Serglycin
2	O84146	Cell fusion glycoprotein K
3	O84251	Membrane protein
4	O84432	Cell division protein anillin
5	O84482	Lipoprotein
6	O84624	Effector from type III secretion system
7	O84625	Effector from type III secretion system

8	O84626	Effector from type III secretion system
9	O84627	Exodeoxyribonuclease V, gamma subunit
10	O84671	Inner membrane component of T3SS*, periplasmic domain
11	O84856	Adenosine specific kinase

*Type three secretion systems

Currently, we are searching for the suitable algorithms that combine the used Bioinformatics tools in the most dynamic way. Once the algorithms are established, an online server will be available for other researchers to delve deeper. Based on predicted functionalities, we have considered that these HPs may play a significant role in the pathogenesis and survival of the C. trachomatis in host organism. It has been suggested that some non-invasive strains of gram-negative bacteria have lost the T3SS because the energetically costly system is no longer functional. Although traditional antibiotics were effective against these bacteria, antibiotic-resistant strains have emerged over the period of time. Understanding the working of T3SS and developing drugs that targets it specifically have become an important goal of many research groups around the world.

3.1 Structural analysis of virulence factors protein

The first step in C. trachomatis infection is assembling the T3SS (Type 3 Secretion System). This bacterial type III secretion system (T3SS) is of special interest to those studying host-pathogen interactions because, by utilizing this system, bacteria are able to directly inject bacterial proteins called effectors into host cells across bacterial and host membranes, where they can manipulate host cell function. So, our study showed 18 HPs may contain some virulence associated factors (Table 1). AlgPred server was used for the prediction of potential allergens in this group of HPs. There were 43 HPs that may act as potential allergens. By utilizing the outcomes of functional and virulence predictions, the 18 HPs were classified as putative virulence factors among the set of 266 functionally characterized HPs (Table 2). The HP O84625 showed highest scores of 1.015, in the cascade search of VirulentPred (Table 3), which was selected for further analyses. The structure of HP O84625 was predicted using ab-initio protocol of I-TASSER. This protein shows all alpha topology (Fig.2.A). In order to observe the.

Table (3): The HPs with highest virulentpred score amongst the T3SS proteins.

S.NO	Uniprot ID	Virulentpred score	Virulentpred	Algpred	DBETH Server	MP3
1	O84624	0.9929(+)	+	+	+	+
2	O84625	1.015(+)	+	+	+	+
3	O84626	0.9389(+)	+	+	+	+

Conformational behavior of HPs with highest virulence scores, the MD simulations were performed at 300 K under explicit solvent conditions. The resulted behaviors of the protein were analyzed and reported in the form of RMSD, Rg, Secondary structure and RMSF (Fig.2). The HP O84625 showed a sharp increase in RMSD values, while sharp fluctuations occur (Fig. 2A). Similarly, the radius of gyration showed steep fluctuations in the values (Fig. 3B) with an average Rg score of 4.0 nm for O84625, that indicates the unstable nature of the proteins. These analyses showed that HP O84625 is highly unstable. The previous work suggested that virulence nature can be associated with the stability of the proteins, as virulent proteins generally showed unstable behavior. The bacterial toxins are produced through a variety of secretory mechanisms present in the bacteria[19]. The disordered regions in the secreted proteins are essential for the virulence[20]. There is no evidence of direct correlation between the structure stability and the virulence of the bacterial proteins. This correlation is only based upon the evidences obtained from MD simulations. Therefore, these HPs can also be considered as the virulence factors and can be utilized for further experimental studies.

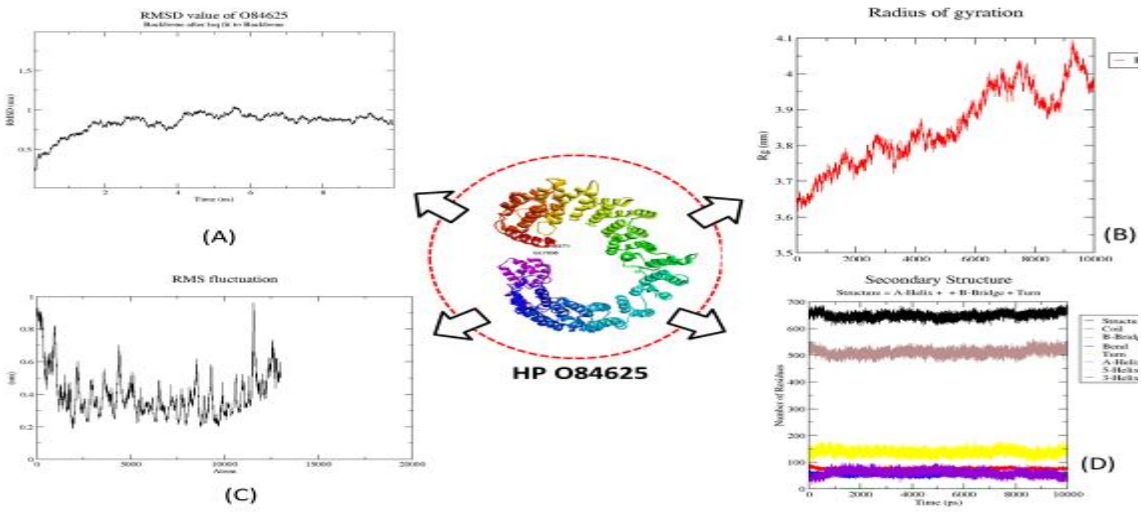

Figure 2. (A) The RMSD value 0.95 during the MD simulation for 10 ns, (B) The plot of radius of gyration (compactness) against the time intervals indicating HP O84625 was unstable in the natural conditions (C), The RMSF plot show sharp fluctuation, (D) It showed secondary structures map.

4. CONCLUSION

Using in silico approach, we successfully examined the sequences of 266 HPs from the genome of Chlamydia trachomatis and predicted the functions and analyzed the virulence characteristics of 11 HPs. Due to the unavailability of the experimental validations and less precision of the available in silico methods, a high degree of uncertainty is present in the virulence and functional predictions of the HPs. These findings are complemented by the characterization of the proteins that may be involved in the signaling as well as secretory pathways. Although, the exact molecular function of an HP cannot be deduced computationally, but our study can provide a framework in order to allocate the probable molecular function by utilizing the sequences of the proteins. In this study, the conformational behaviors of the virulence factors were also analyzed by using the Molecular dynamics techniques, which clarified the dynamical behavior of virulent HPs under explicit water conditions. Our study further facilitates a rapid identification of the HPs that are regarded as potential therapeutic targets and may play a significant role in delving deeper into the host-pathogen interactions. Once these HPs are established as a novel drug/vaccine targets, it can pave way for further research for new inhibitors and vaccines. Since these proteins may play a significant role in the host-pathogen interactions and can be considered as the putative drug targets for the design of better therapeutic agents, such information can be utilized to discover new inhibitor molecules and vaccines for the clinically important organisms.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Laishram Singh, who provided us Bioinformatics facility.

REFERENCES

1. Paavonen, J. and W. Eggert-Kruse, *Chlamydia trachomatis: impact on human reproduction.* Human reproduction update, 1999; 5(5): 433-447.
2. Svensson, L.-O., et al., *Screening for Chlamydia trachomatis infection in women and aspects of the laboratory diagnostics.* Acta obstetricia et gynecologica scandinavica, 1991; 70(7-8): 587-590.
3. Carr, S. and E. Espey, *Intrauterine devices and pelvic inflammatory disease among adolescents.* Journal of Adolescent Health, 2013; 52(4): S22-S28.

4. Patel, A.L., et al., *Prevalence of Chlamydia infection among women visiting a gynaecology outpatient department: evaluation of an in-house PCR assay for detection of Chlamydia trachomatis.* Annals of clinical microbiology and antimicrobials, 2010; 9(1): 1.
5. Malhotra, M., et al., *Prevalence of sexually transmitted infections in patients attending a tertiary care hospital in North India-a retrospective Study.* Indian J Sex Transm Dis, 2008; 29: 82-5.
6. Mittal, A., *Genital chlamydiasis.* ICMR Bull, 1992; 22: 103-6.
7. Vidhani, S., et al., *Seroprevalence of Chalmydia trachomatis infection amongst patients with pelvic inflammatory diseases and infertility.* The Journal of communicable diseases, 2005; 37(3): 233-238.
8. Hillis, S., et al., *New opportunities for Chlamydia prevention: applications of science to public health practice.* Sexually transmitted diseases, 1995; 22(3): 197.
9. Institute, R.R., *The Robinson Research Institute.* 2008.
10. Rosqvist, R., K.-E. Magnusson, and H. Wolf-Watz, *Target cell contact triggers expression and polarized transfer of Yersinia YopE cytotoxin into mammalian cells.* The EMBO journal, 1994; 13(4): 964.
11. Ghosh, P., *Process of protein transport by the type III secretion system.* Microbiology and molecular biology reviews, 2004; 68(4): 771-795.
12. Shahbaaz, M., et al., *In silico approaches for the identification of virulence candidates amongst hypothetical proteins of Mycoplasma pneumoniae 309.* Computational biology and chemistry, 2015; 59: 67-80.
13. Baron, C. and B. Coombes, *Targeting bacterial secretion systems: benefits of disarmament in the microcosm.* Infectious Disorders-Drug Targets (Formerly Current Drug Targets-Infectious Disorders), 2007; 7(1): 19-27.
14. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of molecular biology, 1990; 215(3): 403-410.
15. Finn, R.D., J. Clements, and S.R. Eddy, *HMMER web server: interactive sequence similarity searching.* Nucleic acids research, 2011; gkr367.
16. Studio, D., *Accelrys Inc.* San Diego, CA, USA, 2013.
17. Zhang, Y. and J. Skolnick, *TM-align: a protein structure alignment algorithm based on the TM-score.* Nucleic acids research, 2005; 33(7): 2302-2309.
18. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* Journal of computational chemistry, 2005; 26(16): 1701-1718.

19. McCann, H.C., et al., *Genomic analysis of the kiwifruit pathogen Pseudomonas syringae pv. actinidiae provides insight into the origins of an emergent plant disease.* PLoS Pathog, 2013; 9(7): e1003503.
20. Liu, Y., et al., *Aggregation and morphology control enables multiple cases of high-efficiency polymer solar cells.* Nature communications, 2014; 5.