Mining the Proteome of *Haemophilus ducreyi* for Identification of Potential Drug Targets

Aditya Narayan Sarangi*,[1], Stuti Gupta[1], Nidhi Trivedi[1], Bhawna Rathi[1] and Shailendra Kumar Gupta*,[2]

[1]*Biomedical Informatics Centre, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, India*

[2]*ENVIS Centre, Indian Institute of Toxicology Research, Lucknow, India*

Abstract: Chancroid is an extremely infectious sexually transmitted disease (STD) caused by the bacterium *Haemophilus ducreyi*, prevalent in Africa, United States and in some parts of South Asia. Chancroid has been recognized as a cofactor for human immunodeficiency virus (HIV) transmission. So, there is a requirement to develop an efficient drug to combat chancroid, which can also diminish the HIV prevalence in those populations where chancroid is a prime source for HIV infection. The availability of the complete proteome information of *H. ducreyi* help enabled *in silico* analysis for identification of potential vaccine candidates and drug targets. Our study revealed 1226 proteins in *H. ducreyi* to be non-homologous with human proteome. Screening these proteins using the Database of Essential Genes (DEG) resulted in the identification of 451 essential proteins. Analysis of the identified essential proteins, using the KEGG Automated Annotation Server (KAAS), revealed 40 proteins of *H. ducreyi* as potential drug targets as they are involved in pathogen specific metabolic pathways. Subcellular localization prediction of these 451 essential proteins revealed that 11 proteins lie on the outer membrane of the pathogen which could be potential vaccine candidates. Functional family prediction for the 50 putative uncharacterized essential proteins of *H. ducreyi* by SVM-Prot web server revealed that out of 50, 3 proteins as transmembrane proteins, which may be potential drug targets. Identification of potential inhibitors against these targets through virtual screening may consequence in detection of novel lead compounds effective against *H. ducreyi* to combat chancroid.

Keywords: Proteome mining, Drug target, Vaccine candidate, Chancroid, *Haemophilus ducreyi,* HIV, AIDS.

INTRODUCTION

Chancroid or soft chancre (ulcus molle), is a sexually transmitted disease, caused by facultative anaerobic gram-negative coccobacillus *H. ducreyi*. This disease is common in many tropical and subtropical countries of Africa [1-4] and Southeast Asia [5-7] and has been associated with isolated outbreaks of genital ulcer disease in both North America [8] and Europe [9]. The disease is characterized by painful ulcers of the genitalia which may often be found in association with painful regional lymphadenopathy. The significance of the disease was recently enhanced with the finding that chancroid may be an important cofactor in the heterosexual transmission of human immunodeficiency virus (HIV) [10]. *H. ducreyi* facilitates HIV transmission by providing an accessible portal entry, promoting viral shedding, recruiting macrophages and CD4 cells to the skin [11-15]. The detection and treatment of this disease is urgently needed as it causes the risk of transmission of HIV causing AIDS, which is one of the life threatening diseases.

Centers for Disease Control and Prevention (CDC) recommend azithromycin, ceftriaxone and erythromycin for the treatment of chancroid [16]. However, single-dose treatment by these drugs was found not to be successful in more than 20% of chancroid therapies [17-19]. Furthermore *de novo* mutations and horizontal gene transfer mechanism in *H. ducreyi* may lead to development of drug resistance genes [17, 20]. So, there is a budding need for the identification of novel therapeutic drug targets and drugs against *H. ducreyi* to combat chancroid.

Large amount of genomic and proteomic data of different organisms are available in public domain and various ongoing genomic projects are escalating the genomic data at brisk pace. The completion of the human genome project has revolutionized the field of drug-discovery against life threatening human pathogens [21]. Till date the complete genome sequence of more than one thousand microbes has been determined and another two thousand microbial genome projects are currently in progress [22]. The genomic and proteomic sequence information of pathogens can provide an aid in detection and characterization of the novel therapeutic targets and vaccine candidates.

The subtractive genomic [23] analysis is a widely used *in silico* method that predicts highly conserved genes, which are essential for pathogenic microorganisms and are involved in their replication and survival. These genes are also important components of various metabolic pathways and

*Address correspondence to these authors at the Biomedical Informatics Centre, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow-226014, India; Tel: +91 522 2668004 Extn: 8503; Fax: +91 522 2668017; E-mail: a.n.sarangi@gmail.com and

Indian Institute of Toxicology Research (CSIR), P.O. Box. 80, MG Marg, Lucknow -226001 India; Tel: +91 522 2284591; Fax: +91 522 2628227; E-mail: skgupta@iitr.res.in

mechanisms occurring in the pathogen but do not have any similarity with the host genes. Inhibiting these essential genes which are crucial to sustain cellular life will be fatal for the pathogenic microorganism. This approach significantly narrows down the time required for detection and characterization of potential drug targets.

Genome subtraction approach has been effectively employed to detect novel drug targets in *Pseudomonas aeruginosa* [23]. The current study on *H. ducreyi* is based on proteome subtraction approach. Differential pathway analysis, subcellular localization prediction and functional family classification of putative uncharacterized essential proteins was done to analyze the complete proteome of *H. ducreyi* to find out potential vaccine candidates and therapeutic drug targets for the prevention of chancroid.

METHODOLOGY

Retrieval of Proteomes of Host and Pathogen

The complete proteome of *H. ducreyi* was retrieved from SwissProt [24] and the complete *Homo sapiens* proteome was downloaded from NCBI [25]. The prokaryotic essential proteins were retrieved from the Database of Essential Genes (DEG) [26] (http://tubic.tju.edu.cn/deg) using in-house developed PERL script.

Identification of Essential Proteins in *H. Ducreyi*

The complete proteome of *H. ducreyi* was purged at 60% using CD-HIT [27] to identify the paralogs or duplicate proteins within the proteome of *H. ducreyi*. The paralogs were excluded and the remaining set of protein was subjected to BlastP against complete proteome of *Homo sapiens* with the expectation value (E-value) cut-off of 10^{-4}. The resultant dataset obtained had no significant similarity with proteome of *Homo sapiens*. BlastP analysis was performed for the non-homologous protein sequences of *H. ducreyi* against DEG with E-value cut-off score of 10^{-10}. Minimum bit-score cut-off of 100 was used to filter out only essential proteins, non-homologous to *H. sapiens*, from *H. ducreyi* proteome. An in-house developed PERL script was used to parse the BlastP result.

Metabolic Pathway Analysis

Metabolic pathway analysis of the essential proteins of *H. ducreyi* was done by KEGG Automatic Annotation Server (KAAS) [28]. The metabolic pathways of *Homo sapiens* and *H. ducreyi* were compared to map out essential proteins involved in pathogen specific metabolic pathways only.

Subcellular Localization Prediction

Sub-cellular localization analysis of the essential protein sequences of *H. ducreyi* has been done by Proteome Analyst Specialized Subcellular Localization Server v2.5 (PA-SUB) [29] to identify the surface and membrane associated proteins which could be probable vaccine candidates.

Protein Functional Family Classification

Functional family classification of the putative uncharacterized essential proteins was done by using the SVMProt web server (http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi) [30]. SVMProt utilizes Support Vector Machine (SVM) for classification of a protein into functional family using primary sequence information.

RESULTS & DISCUSSION

The results obtained through subtractive proteomic approach, metabolic pathway analysis and sub cellular localization are listed in Table **1**. The purpose of this analysis was to detect the essential proteins of *H. ducreyi*. The filtered essential proteins of pathogen, non homologous to human, would represent potential drug targets.

The complete proteome set of *H. ducreyi* contains 1694 proteins. 1660 proteins were retrieved as non-redundant by CD-HIT program at 60% threshold. 451 proteins were identified to be essential for the pathogen. These 451 proteins were then further analyzed and their sub-cellular localization was predicted using PA-SUB server to locate the outer membrane proteins which could be probable vaccine candidates. 11 outer membrane proteins were identified, which are listed in the supplementary Table **1**. Membrane proteins have vital role in cellular communications, signal transduction, transport of ions, metabolites and other molecules. These represent a vast number of potential therapeutic drug targets because of their involvement in major biological processes in cell.

Protein functional family prediction provides important information regarding structure, activity and metabolic roles [31]. The essential proteins identified in the pathogen com-

Table 1. **Summary of Subtractive Proteomic and Metabolic pathway Analysis of *H. ducreyi* Proteins**

Summary of Analysis	Result
Total Number of Proteins	1694
Duplicates (>60% identical) in CD-HIT	34
Non-redundant proteins	1660
Non-human homologous proteins (E-value 10^{-4})	1226
Essential protein in DEG (E-value 10^{-10})	451
Essential proteins involved in metabolic pathways	40
Pathways unique to *H. ducreyi*	11
Outer membrane essential proteins of *H. ducreyi*	11

prises of 50 putative uncharacterized proteins. Protein family classification allows probable function assignment for the uncharacterized protein. These 50 putative uncharacterized proteins were characterized and classified by SVM-Prot web server. Among the 50 proteins, 1 protein is classified as DNA repair protein, 5 as DNA binding proteins, 6 as transferases, 1 as ATP binding protein, 8 as lipid binding proteins, 3 as transmembrane proteins, 2 as hydrolases, 1 as zinc binder protein, 2 proteins with oxidoreductive property, 1 as metal binder and 1 as electrochemical potential-driven transporter (Supplementary Table **2**). These transmembrane and transporter proteins predicted by SVM-Prot web server [31] may be considered as effective drug targets.

Information about disease specific target proteins and metabolites involved in different metabolic pathways provides detail insight to discover novel therapeutic targets and help understanding the interaction with other molecules in performing specific tasks [32]. Metabolic pathway analysis of the essential proteins of *H. ducreyi* was done by KEGG Automatic Annotation Server (KASS). The comparative analysis of metabolic pathways of the host and the pathogen revealed 11 pathways that were found to be distinctive in *H. ducreyi*. 40 proteins found to be present exclusively in the pathogen specific metabolic pathways. The results are summarized in Supplementary Table **3**.

Fructose 1, 6-bisphosphate (FBP) aldolase (E.C. 4.1.2.13) plays an important role in catalyzing the reactions involved in calvin cycle, glycolysis and gluconeogenesis. Based on the reaction mechanism, the FBP aldolases are divided in two groups: Class I aldolases forms a schiff-base intermediate with the substrate and Class II aldolases instead use a divalent metal to stabilize the carbanion intermediate. The plants and animals lack Class II aldolases, therefore they can be used as feasible drug targets [32]. FBP aldolases of *H. ducreyi* are found to be involved in many enzymatic pathways, viz. glycolysis, gluconeogenesis, pentose phosphate pathway, fructose and mannose metabolism. Molecular modeling of the derivatives in the enzymes' active site was also used to design a new generation of inhibitors [33].

Acetyl-coenzyme A Carboxylases (ACCs) are responsible for fatty acid metabolism in most living organisms. So, they are proved as possible targets for drug discovery against a variety of human diseases, including diabetes, obesity, cancer and microbial infections [34]. It can also be considered as a potential therapeutic target against chancroid disease as it is found to be essential in *H. ducreyi* but not in human.

The Phosphoenolpyruvate Carbohydrate Phosphotransferase system (PTS) is confined exclusively in bacteria, and plays a major role in catalyzing the transport and phosphorylation of numerous monosaccharides, disaccharides, amino sugars, polyols and other sugar derivatives [35]. The *H. ducreyi* PTS contains five proteins, in which glucose-specific PTS system component and mannose-specific phosphotransferase IIAB component shows activity in multiple metabolic pathways like glycolysis, gluconeogenesis, starch and sucrose metabolism, amino sugar and nucleotide sugar metabolism and fructose and mannose metabolism. Therefore these two proteins can be effectively analyzed and can also be considered as potential drug targets.

Beginning from *in silico* identification of putative drug targets to their validation as potential drug targets is an intricate challenge [36]. Several promising technologies for identification and validation of protein targets include two-dimensional gel electrophoresis, affinity purification, mass spectrometry, protein arrays, isotope-encoding, two-hybrid systems, information technology and activity-based assays [37]. These proteomic techniques can be applied to drug discovery pipeline for identification and validation of potential drug targets, drug-protein interaction and drug specificity.

CONCLUSION

The large scale genome sequencing projects have increased the availability of completely sequenced genomic and proteomic data in public domain. Screening and analysis of these large biological sequence data provide new opportunities to understand and combat both infectious and genetic diseases in humans. There is a budding need for new drugs and vaccines to treat and prevent emerging and neglected infectious diseases. Subtractive genomics is a powerful tool for exploring new therapeutic targets. The current study based on subtractive proteomics approach helped in the identification and characterization of the potential essential proteins that could be targets for efficient drug designing against *H. ducreyi* to combat chancroid. Furthermore, molecular modeling of the drug targets will decipher the best possible active sites that can be targeted by simulations for drug design. Screening these potential targets against drug bank might be useful in the discovery of potential therapeutic compounds against *H. ducreyi*.

ABBREVIATIONS

STD	=	Sexually Transmitted Disease
HIV	=	Human Immunodeficiency Virus
AIDS	=	Acquired Immune Deficiency Syndrome
CDC	=	Centers for Disease Control and Prevention
DEG	=	Database of Essential Genes
KEGG	=	Kyoto Encyclopedia of Genes and Genomes
KAAS	=	KEGG Automated Annotation Server
SVM	=	Support Vector Machine
FBP	=	Fructose 1, 6-bisphosphate
ACCs	=	Acetyl-coenzyme A carboxylases
PTS	=	Phosphotransferase System.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publishers Web site along with the published article.

REFERENCES

[1]   M. O. Duncan, Y. R. Bilgeri, H. G. Fehier and R. C. Ballard, "The diagnosis of sexually acquired genital ulcerations in black patients

in Johannesburg. S. Afr.", *J. Sex. Transm. Dis.*, vol. 1, pp. 20-23, 1981.

[2] A. S. Latif, "Sexually transmitted disease in clinic patients in Salisbury, Zimbabwe", *Br. J. Vener. Dis.*, vol. 57, pp. 81-183, 1981.

[3] A. Meheus, E. van Dyck, J. P. Ursi, R. C. BaUlard and P. Piot, "Etiology of genital ulcerations in Swaziland", *Sex. Transm. Dis.*, vol. 10, pp. 33-35, 1983.

[4] H. Nsanze, M. V. Fast, L. J. D'Costa, P. Tukei, J. Curran and A. R. Ronald, "Genital ulcers in Kenya: a clinical and laboratory study", *Br. J. Vener. Dis.*, vol. 57, pp. 378-381, 1981.

[5] V. S. Rajan, E. H. Sng and A. L. Lin, "The isolation of *H. ducreyi* in Singapore", *Ann. Acad. Med. Singapore*, vol. 12, pp. 57-60, 1983.

[6] U. Tan, V. S. Rajan, S. L. Koe, N. J. Tan, B. H. Tan and A. J. Goh, "Chancroid: a study of 500 cases", *Asian J. Infect. Dis.*, vol. 1, pp. 27-28, 1977.

[7] D. N Taylor, C. Pitarangsi, P. Echeverria, K. Panikabutra and C. Suvongse, "Comparative study of ceftriaxone and trimethoprim-sulfamethoxazole for the treatment of chancroid in Thailand", *J. Infect. Dis.*, vol. 152, pp. 1002-1006, 1985.

[8] G. P. Schmid, "The treatment of chancroid", *J. Am. Med. Assoc.*, vol. 255, pp.1757-1762, 1986.

[9] K. C. Nayyar, E. Stolz and M. F. Michel, "Rising incidence of chancroid in Rotterdam: epidemiological, clinical, diagnostic and therapeutic aspects", *Br. J. Vener. Dis.*, vol. 55, pp. 439- 441, 1979.

[10] J. K. Kreiss, D. Koech, F.A. Plummer, K. K. Holmes, M. Lightfoote, P. Piot, A. R. Ronald, J. O. Ndinya-Achola, L. J. D'Costa, P. Roberts, E. N. Ngugi and T.C. Quinn, "AIDS virus infection in Nairobi prostitutes. Spread of the epidemic to East Africa", *N. Engl. J. Med.,* vol. 314, pp. 414-418, 1986.

[11] P. G. Jessamine and A. R. Ronald, "Chancroid and the role of genital ulcer disease in the spread of human retrovirus", *Med. Clin. North Am.*, vol. 74, pp. 1417-1431, 1990.

[12] J. N. Wasserheit, "Interrelationships between human immunodeficiency virus infection and other sexually transmitted diseases", *Sex Transm.*, vol. 19, pp. 61-77, 1992.

[13] S. M. Spinola, A. Orazi, J. N. Arno, K. Fortney, P. Kotylo, C. Y. Chen, A. A. Campagnari and A. F. Hood, "Haemophilus ducreyi elicits a cutaneous infiltrate of CD4 cells during experimental human infection", *J. Infect. Dis.*, vol. 173, pp. 394-402, 1996.

[14] R. King, J. Gough, A. Ronald, J. Nasio, J. O. Ndinya-Achola, F. Plummer and J. A. Wilkins, "An immunohistochemical analysis of naturally occurring chancroid", *J. Infect. Dis.*, vol. 174, pp. 427-430, 1996.

[15] J. K. Kreiss, R. Coombs, F. Plummer, K. K. Holmes, B. Nikora, W. Cameron, E. Ngugi, J. O. N. Achola and L. Core, " Isolation of human immunodeficiency virus from genital ulcers in Nairobi prostitute", *J. Infect. Dis.*, vol. 160, pp. 380-384, 1989.

[16] D. L. Trees and S. A. Morse, "Chancroid and Haemophilus ducreyi: an update", *Clin. Microbiol. Rev.*, vol. 8, pp. 357-375,1995.

[17] J. E. Leon, W. D. Lauzon, B. Toye, N. Singhal and D. W. Cameron, "*In vitro* and *in vivo* activity of combination antimicrobial agents on Haemophilus ducreyi", *J. Antimicrob. Chemother.*, vol. 56, no. 3, pp. 552-558, 2005.

[18] K. S. MacDonald, D. W. Cameron, L. D'Costa, J. O. Ndinya-Achola, F. A. Plummer and A. R. Ronald, " Evaluation of fleroxacin (RO 23-6240) as single-oral-dose therapy of culture-proven chancroid in Nairobi, Kenya", *Antimicrob. Agents Chemother.*, vol. 33, pp. 612-614, 1989.

[19] M. W. Tyndall, E. Agoki, F. A. Plummer, W. Malisa, J. O. Ndinya-Achola and A. R. Ronald, "Single dose azithromycin for the treatment of chancroid: a randomized comparison with erythromycin", *Sex. Transm. Dis.,* vol. 21, pp. 231-234, 1994.

[20] P. J. Willson, W. L. Albritton, L. Slaney and J. K. Setlow, "Characterization of a multiple antibiotic resistance plasmid from Haemophilus ducreyi", *Antimicrob. Agents Chemother.,* vol. 33, pp. 1627-1630, 1989.

[21] M. Y. Galperin, and E. V. Koonin, "Searching for drug targets in microbial genomes", *Curr. Opin. Biotechnol.*, vol. 10, pp. 571-578, 1999.

[22] Complete Microbial Genomes. [Online]. Available: http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi [Accessed: Nov 16, 2009].

[23] K. R. Sakharkar, M. K. Sakharkar and V. T. Chow, "A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa", *In Silico Biol.*, vol. 4, pp. 355-360, 2004.

[24] A. Bairoch and R. Apweiler, "The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000", *Nucleic Acids Res.*, vol. 28, pp. 45-48, 2000.

[25] D. Maglott, J. Ostell, K. D. Pruitt and T. Tatsuova, "Entrez Gene: gene-centered information at NCBI", *Nucleic Acids Res.,* vol. 35: pp. D26-D31, 2007.

[26] R. Zhang and Y. Lin, "DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes", *Nucleic Acids Res.*, vol. 37 (Database issue), pp. D455-D458, 2009.

[27] W. Li and A. Godzik, "Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences", *Bioinformatics*, vol. 22, pp. 1658-1659, 2006.

[28] Y. Moriya, M. Itoh, S. Okuda, A. C. Yoshizawa and M. Kanehisa, "KAAS: an automatic genome annotation and pathway reconstruction server", *Nucleic Acids Res.*, vol. 35 (Web Server issue), pp. W182-185, 2007.

[29] Z. Lu, D. Szafron, R.Greiner, P. Lu, D. S. Wishart, B. Poulin, J. Anvik, C. Macdonell and R. Eisner, "Predicting subcellular localization of proteins using machine-learned classifiers", *Bioinformatics*, vol. 20, pp. 547-556, 2004.

[30] C.Z. Cai, L.Y. Han, Z. L. Ji, X. Chen and Y.Z. Chen, "SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence". *Nucleic Acids Res.*, vol. 31, pp. 3692-3697, 2003.

[31] H. W. Cathy, H. Huang, L. L. Yeh and W. C. Barker, "Protein family classification and functional annotation", *Comput. Biol. Chem.*, vol. 27, pp. 37-47, 2003.

[32] C. J. Zheng, H. Zhou, B. Xie, L. Y. Han, C. W. Yap and Y. Z. Chen, "TRMP: a database of therapeutically relevant multiple pathways", *Bioinformatics*, vol. 20, pp. 2236-2241, 2004.

[33] G. Labbé, A. Krismanich, S. de Groot, T. Rasmusson, G. I. Dmitrienko and J. Guy, "Guillemette Metal binding and inhibition of Class II fructose 1,6-bisphosphate aldolase", *FASEB J.*, vol. 21, p. 804, 2007.

[34] L. Tong and H. J. Jr Harwood., "Acetyl-coenzyme A carboxylases: versatile targets for drug discovery", *J. Cell Biochem.*, vol. 99, pp. 1476-1488, 2006.

[35] J. Deutscher, C. Francke and P. W. Postma, "How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria", *Microbiol. Mol. Biol. Rev.*, vol. 70, pp. 939-1031, 2006.

[36] K.K. Kopec, D. Bozyczko-Coyne and M. Williams, "Target identification and validation in drug discovery: the role of proteomics", *Biochem. Pharmacol.*, vol. 69, pp. 1133-1139, 2005.

[37] J. Burbaum and G.M.Tobal, "Proteomics in drug discovery", *Curr. Opin. Chem. Biol.,* vol. 6, pp. 427-433, 2002.