Bioinformation
www.bioinformation.net

open access
Hypothesis

Identification and modeling of a drug target for *Clostridium perfringens* SM101

Gagan Chhabra[#], Pramila Sharma[#], Avishek Anant, Sachin Deshmukh, Himani Kaushik, Keshav Gopal, Nutan Srivastava, Neeraj Sharma, Lalit C. Garg*

Gene Regulation Laboratory, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi – 110067, India; [#]Equal contribution; Dr. Lalit C. Garg - E-mail : lalit@nii.res.in; Tel.: +91 11 26703652; Fax : +91 11 26742125; *Corresponding author

Received December 14, 2009; Revised December 26, 2009; Accepted December 28, 2009; Published January 17, 2010

Abstract:
In the present study, comparative genome analysis between *Clostridium perfringens* and the human genome was carried out to identify genes that are essential for the pathogen's survival, and non-homologous to the genes of human host, that can be used as potential drug targets. The study resulted in the identification of 426 such genes. The number of these potential drug targets thus identified is significantly lower than the genome's protein coding capacity (2558 protein coding genes). The 426 genes of *C. perfringens* were further analyzed for overall similarities with the essential genes of 14 different bacterial species present in Database of Essential Genes (DEG). Our results show that there are only 5 essential genes of *C. perfringens* that exhibit similarity with 12 species of the 14 different bacterial species present in DEG database. Of these, 1 gene was similar in 12 species and 4 genes were similar in 11 species. Thus, the study opens a new avenue for the development of potential drugs against the highly pathogenic bacterium. Further, by selecting these essential genes of *C. perfringens*, which are common and essential for other pathogenic microbial species, a broad spectrum anti-microbial drug can be developed. As a case study, we have built a homology model of one of the potential drug targets, ABC transporter-ATP binding protein, which can be employed for *in silico* docking studies by suitable inhibitors.

Keywords: *Clostridium perfringens,* DEG, Essential genes, Drug targets, Broad-spectrum anti microbial drug.

Background:
The availability of the complete genome sequence information of the human genome and a large number of microbial genomes' sequences has led to the development of new approaches to understand host-pathogen interaction. Use of bioinformatics approach and comparative analysis of the genome of a pathogenic microbe allows one to identify essential genes necessary for the survival of that pathogen. The proteins encoded by these essential genes, that are not present or are non-homologous to the host, can be used as drug targets. Such an approach has been effectively used to identify drug targets in other bacterial species such as *Pseudomonas aeruginosa* **[1, 2],** *Helicobacter pylori* **[3],** *Mycobacterium tuberculosis* **[4],** *Burkholderia pseudomallei* **[5]** and *Aeromonas hydrophila* **[6].** *Clostridium perfringens* is a Gram-positive, rod shaped, anaerobic bacterium that is able to form spores. It is widely distributed in the environment (e.g. in soil and sewage) and is frequently found in the gastrointestinal (GI) tract of humans, many domestic and feral animals, as well as in soil and freshwater sediments **[7].** In humans, it can cause gangrene and gastrointestinal disease (e.g. food poisoning and necrotic enteritis), whereas in other animals, gastrointestinal and enterotoxemic diseases occur more frequently **[8].** *C. perfringens* does not invade healthy cells but produces various toxins and enzymes that are responsible for associated lesions and symptoms. As a species, *C. perfringens* is one of the most prolific producers of toxins **[9].** It has five biotypes (A, B, C, D and E), which are identified by the main types of toxins they produce (alpha, beta, iota, epsilon and theta), each type of toxin being associated with a specific syndrome. *C. perfringens* type A is the most common toxin type in the environment, and is responsible for gas gangrene, enterocolitis, dysenteria, and enterotoxemia. In the present study, comparative genome analysis of *C. perfringens* type A with that of the human genome, and use of the Database of Essential Genes (DEG) compiled by Zhang *et al.*, **[10],** have resulted in the identification of the essential genes of *C. perfringens,* that could be used as potential drug targets.

Methodology:
Comparative genome analysis:
The complete genomes of *C. perfringens type A,* strain SM101 (Accession No. NC_008262) **[11]** and its human host have been sequenced and were downloaded from the NCBI website **[12].** The Database of Essential Genes **[10]** was accessed from its website **[13]** and sequence alignment was performed using BLASTP. In this analysis, the assumption described by Dutta *et al.*, **[3]** was followed, and proteins of less then 100 amino acid residues were excluded from the analysis. The remaining proteins were subjected to BLASTP on the NCBI server against human protein sequences to identify non-homologous sequences **[14].** A minimum bit score of 100 and an Expectation value (E-value) cutoff of 10^{-10} was selected for shortlisting genes. The shortlisted genes were subjected to BLASTP against DEG to identify essential genes. Further analysis of the essential genes using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway database **[15],** revealed the information about different biological process in which potential target genes were involved.

Structure modeling and visualization of model:
BLASTP analysis was used to identify the most suitable template for homology modeling of *Clostridium perfringens* ABC transporter, ATP binding protein (CpABC) (Accession No. YP_698054). Subsequent to BLASTP analysis, the sequences of the structures of ABC transporters available in the PDB were used. The available structure of ABC transporter from *Methanococcus jannaschii* (Mj0796) in the Protein Database (PDB entry 1f3o, resolution=2.70, R value=0.204) was used as a template. The target and the template sequences were aligned using ClustalW. MODELLER **[16],** an automated comparative protein modeling program, was used for homology modeling to generate the 3-D structure of CpABC. Further, the structural model generated was visualized using the Swiss PDB viewer software **[17].**

Validation of the generated model:
Different structure verification servers such as PROCHECK **[18],** WHAT_CHECK **[19],** VERIFY3D **[20]** and ProSA **[21]** were used to evaluate the 3D-model. These verification programs validate the predicted structure by checking various parameters. While PROCHECK, a structure verification program that relies on Ramachandran plot **[22],** determines the quality of the predicted structure by assessing various parameters such as lengths, angles and planarity of the peptide bonds, geometry of the hydrogen bonds, and side chain conformations of protein structures as a function of atomic resolution, WHAT_CHECK, a subset of protein verification tools from the WHATIF **[23],** program extensively evaluates the stereochemical parameters of the residues in the model **[24].** The Verify3D determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D) by assigning a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar etc.) and comparing the results to valid structures **[25].**

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 278-289 (2010)

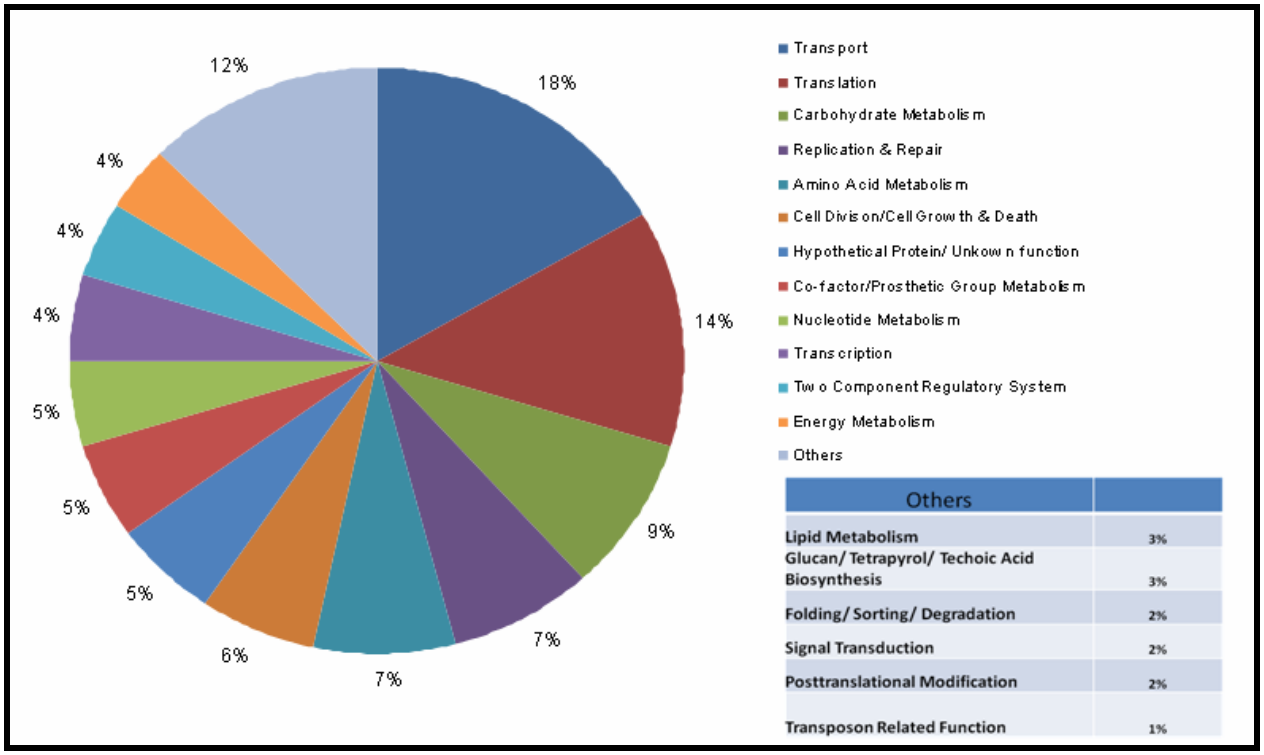

Figure 1: Percentage distribution of 426 target genes encoding different classes of proteins in *Clostridium perfringens.* Around 2% gene encode the proteins of unknown function.

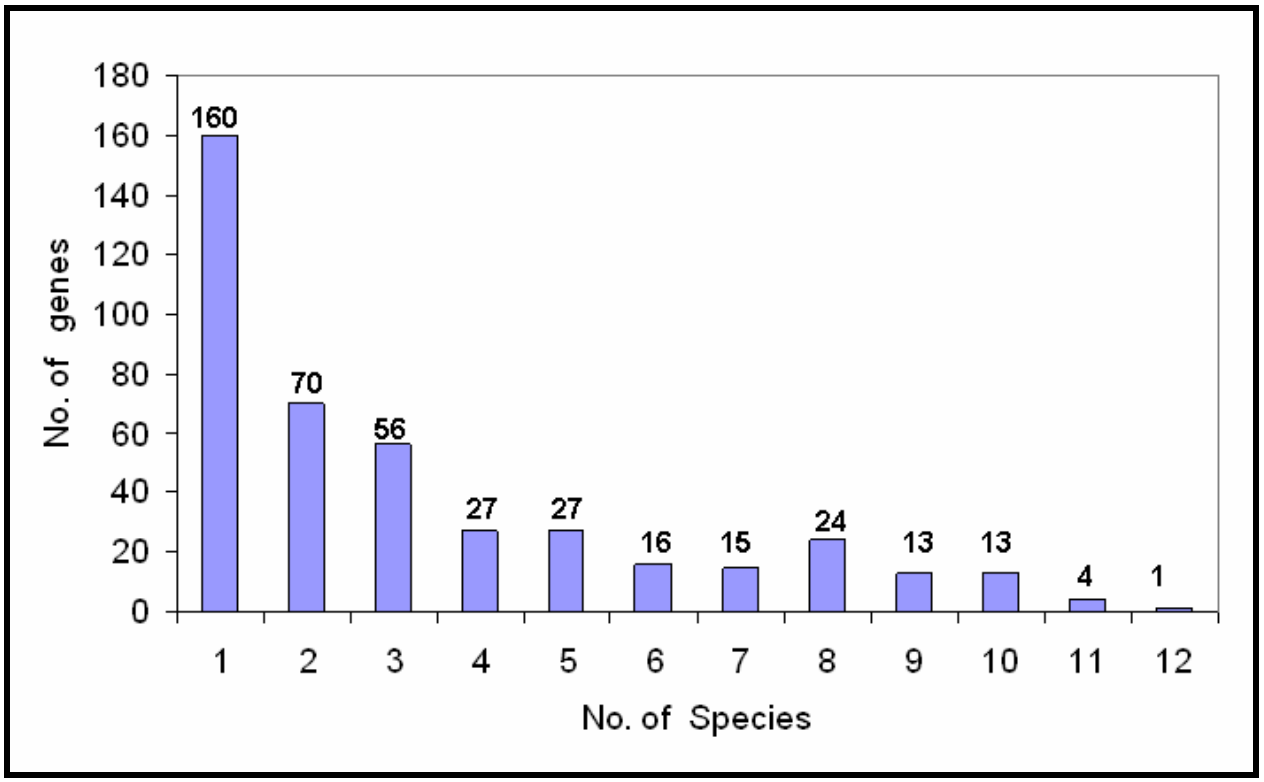

Figure 2: Graphical representation of the number of target genes aving similarity with those of bacterial species present in DEG. The number of genes of *C. perfringens* having similar match to different number of bacterial species is shown on top of the respective bars.

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 278-289 (2010)

Figure 3: CLUSTALW Multiple sequence alignment of CpABC with Mj0796. Single fully conserved residues are represented by (*), conservation of strong and weak groups is denoted by (:) and (.), respectively. The boxed sequence represents the Walker A motif, whereas the ABC signature sequence is marked by bold overline.

Discussion:

Identification of drug targets in *C. perfringens*:

C. perfringens is the most common cause of gas gangrene in humans. Clostridial gas gangrene is a highly lethal necrotizing soft tissue infection of skeletal muscle caused by toxins secreted by *C. perfringens.* Although penicillin is one of the preferred antibiotics, it is only useful if the infection is diagnosed at early stages. There is no other specific drug that can be given to a patient infected with *C. perfringens*. Therefore, more research in this field is required to identify new drug targets and develop therapeutic agents for controlling *C. perfringens* infections. Since most antibiotics target essential cellular processes, essential gene products of microbial cells are promising new targets for antibacterial drugs **[10].** Targetting an essential gene necessary for the bacterial cell survival may provide an effective way to control infection.

The circular genome of *C. perfringens* comprises 2,897,393 nucleotides with a total number of 2701 genes. Of the 2558 protein encoding genes, only 2300 genes encode proteins of greater than 100 amino acid residues. BLAST analysis of these genes against the human genome sequence revealed 1991 genes to be non-homologous to humans. Further BLASTP analysis of the 2300 protein coding genes with DEG resulted in identification of 726 genes, which had a bit score of at least 100 at an expectation cutoff value of 10^{-10}, as similar to the essential genes required for the growth and survival of bacteria listed in the DEG. Of these, 426 were found to have no human homologue (see **Table 1** in Supplementary data). Pathways information for these genes was obtained from KEGG database. All these genes are involved in the production of proteins that are useful for various important functions in *C. perfringens.* Out of the 426 identified genes, function of 10 genes remains unknown, and 17 genes encode conserved hypothetical proteins. The percentage distribution of the genes amongst different biological process is shown in **Figure 1.** A large population of these genes (~33%) is involved in metabolic pathways. The major share of these genes constitute the proteins involved in transport and translation (17% and 12%, respectively). Highly conserved genes, in theory, are more likely to be physiologically important **[26];** however, they need to be experimentally validated. Therefore, the analysis of the 426 essential genes of *C. perfringens* for overall similarities with all 14 species present in DEG database was carried out. Results of such an analysis are shown in **Figure 2.** Out of the 426 essential genes, 160 genes have similar match to at least 1 species, whereas on the other end of the spectrum, only 4 genes have similar match to 11 species and only 1 gene has an identity score of more than 100 with 12 different microbial species listed in the DEG. From this analysis, it is evident that the products of 5 genes (1 gene similar in 12 species and 4 genes similar in 11 species) are essential for most of the bacterial species present in DEG. These species include *Acinetobacter baylyi* ADP1, *Bacillus subtilis*, *Escherichia coli* MG1655, *Francisella novicida* U112, *Haemophilus influenzae, Helicobacter pylori* 26695, *Helicobacter pylori* J99**,** *Mycobacterium tuberculosis* H37Rv, *Mycoplasma genitalium* G37, *Mycoplasma pulmonis* UAB CTIP , *Salmonella typhimurium, Staphylococcus aureus* and *Streptococcus pneumoniae.* Therefore, these 5 genes can be used as potential drug targets for more than 10 highly pathogenic bacterial species, in addition to *C. perfringens*. These 5 target genes, thus identified, are ABC transporter-ATP-binding protein, cell division protein FtsZ, RNA polymerase sigma factor RpoD, 50S ribosomal protein L13, and 30S ribosomal protein S5. A drug designed against these targets can be effectively used to treat other bacterial infections as well. Since the number of thus identified potential candidate genes is relatively small, these can be experimentally validated to develop broad-spectrum antimicrobial drugs. Since ABC transporter-ATP-binding protein is one of the 5 potential drug targets identified, an attempt has been made to predict its structure for effective drug design.

ABC transporter - a potential broad spectrum target:

ABC (ATP-binding cassette) transporters are ubiquitously present ATP-dependent transmembrane solute pumps and ion channels. These superfamilies contain both uptake and efflux transport systems and form one of the largest transporters **[27].** The ABC transporters couple hydrolysis of ATP to the translocation of various substrates across cell membranes. Members of this superfamily recognize substrates ranging from single ions to entire protein toxins. ABC transporters have a number of highly conserved ABC cassette motifs, many of which are involved in the binding and hydrolysis of ATP. It is generally assumed that all ABC cassettes bind and hydrolyze ATP in a similar way and use a common mechanism to provide energy for substrate transport through the membrane-spanning domains **[28].** When the substrate has traversed the membrane, the transporter returns to the resting state through dissociation of ADP and inorganic phosphate. Fluoxetine and omeprazole, few of the most widely prescribed drugs in the world, have a transporter protein as site of action. Therefore, ABC transporter structures have potential value in drug designing.

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 278-289 (2010)

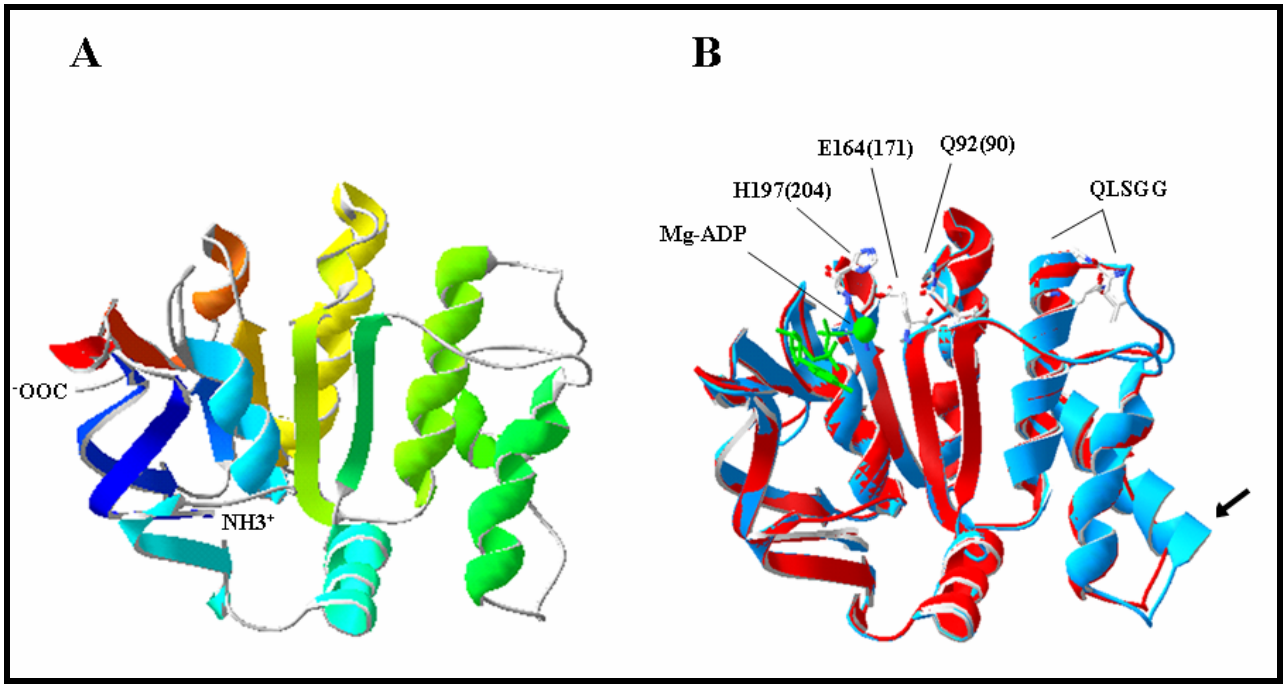

Figure 4: (A): Homology modeled structure of the *C. perfringens* ABC transporter, ATP binding protein based on template Mj0796. Model is represented in ribbon form as Swiss PDB viewer representation in secondary structure succession color scheme. N and C termini of modeled structure are represented as $NH3^+$ and ^-OOC. **(B)** Superimposed image of the modeled structure of CpABC onto Mg-ADP bound Mj0796 (PDB entry 1f3o). CpABC is represented in red color, Mg-ADP (shown in green color) bound Mj0796 is represented in blue color. Signature sequence LSGG with three conserved residues, Q90, E171, H204 form Mj0796 and Q92, E164, H197 from model CpABC are mentioned. Arrow indicates the deletion of seven amino acids (RKRALEC) in CpABC, which forms an α-helix in the Mj0796.

Structural model and overall architecture of ABC transporter (CpABC):

Sequence alignment of the *C. perfringens* CpABC and ABC transporters from other species revealed Mj0796 to be the best template for homology modeling of the target sequence as the CpABC and Mj0796 shared 41% identity **(Figure 3).** Mj0796 is a member of the o228/LolD transporter family, involved in the export of lipoproteins via the Lol system. LolD contains a characteristic sequence called the LolD motif, which is highly conserved among LolD homologs, but not in other ABC transporters, and is located between the Walker A (GSGKST, boxed) and ABC signature (LSGGQ, marked as bold overline) motifs **(Figure 3).** Comparative sequence analyses, motif search, and secondary structure prediction indicated that CpABC is structurally similar to the monomer structure of the Mj0796, a lipid transporter. The crystal structure of Mj0796 (PDB entry 1f3o) was used as a template to predict the structure of CpABC and the predicted 3D structure model of CpABC was generated by Modeller, a homology-modeling program. **Figure 4** shows the predicted structure in the form of ribbons as a Swiss PDB viewer representation.

Validation of generated model of CpABC:

The quality of the model was evaluated using the PROCHECK program and assessed using the Ramachandran plot. It is evident from the Ramchandran plot that the predicted model has 91.4%, 8.1%, and 0.5% residues in the most favorable regions, the allowed regions, and the disallowed regions, respectively. Such a percentage distribution of the protein residues determined by Ramachandran plot shows that the predicted model is of good quality. All Ramachandrans show 6 labelled residues out of 220 whereas chi1-chi2 plots show 2 labelled residues out of 140. The model shows all the main chain and side chain parameters to be in the 'better' region. Another factor that is important for the predicted model to be reliable is G-factor, which is a log odds score based on the observed distribution of stereochemical parameters. For a reliable model, the score for G-factor should be above -0.50. The observed G-factor score for the present model was -0.05 for dihedrals bonds, -0.31 for covalent bonds, and -0.15 overall. The distribution of the main chain bond lengths and bond angles were 98.5% and 93.2% within limits, respectively. Also, all the planar groups were within the limits. The quality of the generated model of CpABC as evaluated by ProSA provided a z-score of -7.2, which falls within the range of values observed for the experimentally determined structures of similar lengths. The validity of the predicted model of CpABC was also verified by employing the structure verification servers WHAT_CHECK and Verify-3D. Superposition of the predicted structure of CpABC and the Mg-ADP bound Mj0796 (template, PDB entry 1f3o) is shown in **Figure 4B.** It is evident from the figure that the Mg-ADP binding core of the ABC subunit and all the structural motifs are highly conserved in both structures. The two superposed structures match 214 Cα atoms with an rms distance of 0.47Å. Three residues Q90, E171, and H204, important for activity of Mj0796, superposed very well with conserved residues Q92, E164 and H197 from model CpABC. However, a deletion of seven amino acids (RKRALEC), which forms an α-helix in the Mj0796 (indicated by arrow), and an insertion of three amino acids (PIS) at the C-terminal end of CpABC, is observed. Thus, the predicted model structure of *C. perfringens* ABC transporter, ATP binding protein, and CpABC is comparable to the structurally resolved Mj0796.

Conclusion:

Comparative genome analysis is a highly efficient approach for idendifing potential proteins that can be used as potential targets for effective drug designing against pathogenic organisms. It allows restricting the potential pool of genes to a much smaller number, compared to the whole genome capacity, which can be experimentally validated. In the present study, around 426 drug targets in *C. perfringens* were identified by comparative genome analysis with DEG.

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 278-289 (2010)

Further, by using the subtractive genomic approach five essential genes were identified that are conserved in more than 10 other pathogenic organisms. Since the number of these conserved genes is very small, these can be experimentally tested for the development of a broad-spectrum anti-microbial drug. The drug thus developed is likely to inhibit other bacterial infections, which share high sequence similarity with the five essential genes of *C. perfringens* SM101.

Acknowledgements:
Financial assistance from the Department of Biotechnology, New Delhi, India is acknowledged. The Council of Scientific and Industrial Research, New Delhi, India, and the University Grants Commission, New Delhi, India is gratefully acknowledged for senior research fellowships to AA and KG, respectively.

References:

[1] KR Sakharkar *et al., In Silico Biology* **4**:355 (2004) [PMID: 15724285]
[2] D Perumal *et al.*, In Silico Biology **7**:453 (2007) [PMID: 18391237]
[3] A Dutta *et al.*, *In Silico Biology* **6:**43 (2006) [PMID: 16789912]
[4] S Anishetty *et al.*, *Computational Biology and Chemistry* **29**:368 (2005) [PMID: 16213791]
[5] CE Chong *et al., In Silico Biology* **6**:341(2006) [PMID: 16922696]
[6] V Sharma *et al., In Silico Biology* **8**:331 (2008) [PMID: 19032165]
[7] JI Rood, ST Cole, *Microbiology Review* **55**:621 (1991) [PMID: 1779929]
[8] JG Songer, Clin. *Microbiology Review* **9:**216-234 (1996) [PMID: 8964036]
[9] JI Rood, *Annual Review of Microbiology* **52**:333 (1998) [PMID: 9891801]
[10] R Zhang *et al., Nucleic Acids Research* **32**:D271 (2004) [PMID:14681410]
[11] GS Myers *et al.,* Genome Research **16:**1031 (2006) [PMID: 16825665]
[12] http://www.ncbi.nlm.nih.gov/
[13] http://tubic.tju.edu.cn/deg/
[14] SF Altschul *et al.*, *Journal of Molecular Biology* **215**:403 (1990) [PMID: 2231712]
[15] H Ogata *et al., Nucleic Acids Research* **27**:29 (1999) [PMID: 9847135]
[16] A Fiser, A Sali, *Methods in Enzymology* **374:**461 (2003) [PMID: 14696385]
[17] http://www.expasy.org/spdbv
[18] RA Laskowski *et al., Journal of Applied Crystallography* **26**:283 (1993)
[19] RWW Hooft *et al., Nature* **381**:272 (1996) [PMID: 8692262]
[20] D Eisenberg *et al., Enzymology* **277**:396 (1997) [PMID: 9379925]
[21] M Wiederstein, MJ Sippl, *Nucleic Acids Research*, **35** (Web Server issue), W407 (2007) [PMID: 17517781]
[22] GN Ramachandran, V Sasisekharan, *Advances in Protein Chemistry* **23**:283 (1968) [PMID: 4882249]
[23] G Vriend, *Journal of Molecular Graphics* **8**:52 (1990) [PMID: 2268628]
[24] http://swift.cmbi.ru.nl/gv/whatcheck/
[25] http://nihserver.mbi.ucla.edu/Verify_3D/
[26] F Arigoni, *Nature Biotechnology* **16**:851 (1998) [PMID: 9743119]
[27] MH Saier, Jr, *Microbiology* **146:**1775 (2000) [PMID: 10931885]
[28] KP Locher *et al., Science* **296:**1091 (2002) [PMID: 12004122]

Edited by P. Kangueane
Citation: Chhabra *et al.*, Bioinformation 4(7): 278-289 (2010)

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 278-289 (2010)

Supplementary material

Table 1: List of genes that are non-homologs to humans and essential for *C. perfringens*

S. No.	Name
1.	chromosomal replication initiator protein DnaA
2.	DNA polymerase III, beta subunit
3.	DNA replication and repair protein RecF
4.	DNA gyrase, A subunit
5.	cytidine/deoxycytidylate deaminase family protein
6.	MTA/SAH nucleosidase
7.	HesA/MoeB/ThiF family protein
8.	glucose-1-phosphate adenylyltransferase
9.	SACPA operon antiterminator (sacT)
10.	pts system, glucose-specific iia component
11.	PTSsystem,N-acetylglucosamine-specific IIBC component
12.	DNA-binding response regulator
13.	permease, putative domain protein
14.	hypothetical protein (CPR_0125)
15.	NAD-dependent malic enzyme
16.	carbamate kinase
17.	N-acetylmannosamine-6-phosphate 2-epimerase, putative
18.	cobalt ABC transporter, ATP-binding protein
19.	acetate kinase
20.	ferrichrome ABC transporter (permease)-like protein lin2287
21.	ferrichrome ABC transporter
22.	hypothetical protein (CPR_0223)
23.	DNA-binding response regulator
24.	GGDEF/EAL domain-containing protein
25.	capsular polysaccharide synthesis protein
26.	teichoic acid translocation permease protein
27.	teichoic acid ABC transporter, ATP-binding protein
28.	tetrapyrrole methylase family protein
29.	Single-strand binding protein family
30.	carboxyl-terminal protease
31.	excinuclease ABC, B subunit
32.	MutS domain protein
33.	L-lactate permease
34.	electron transfer flavoprotein, beta subunit
35.	tldD protein truncated
36.	TldD/PmbA family protein
37.	excinuclease ABC, A subunit
38.	cell cycle protein, FtsW/RodA/SpoVE family
39.	Penicillin binding protein transpeptidase domain protein
40.	excinuclease ABC, C subunit
41.	UDP-N-acetylenolpyruvoylglucosamine reductase
42.	ABC transporter, permease protein, FecCD family
43.	iron compound ABC transporter, ATP-binding protein, putative
44.	3-oxoacyl-(acyl-carrier-protein) synthase III
45.	acetyl-CoA carboxylase, biotin carboxyl carrier protein
46.	beta-hydroxyacyl-(acyl-carrier-protein) dehydratase FabZ
47.	acetyl-CoA carboxylase, carboxyl transferase, beta subunit
48.	acetyl-CoA carboxylase, carboxyl transferase, alpha subunit
49.	ribosomal large subunit pseudouridine synthase B
50.	probable flavoprotein, YhiN family
51.	cytidylate kinase
52.	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
53.	citrate lyase, alpha subunit
54.	malate oxidoreductase (NAD) (malic enzyme)
55.	formate acetyltransferase
56.	pyruvate formate-lyase activating enzyme
57.	LexA repressor
58.	site-specific recombinase, phage integrase family
59.	lysine-specific permease
60.	undecaprenol kinase, putative
61.	mannose-6-phosphate isomerase, class I
62.	Mur ligase family protein

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 278-289 (2010)

63. CobB/CobQ family glutamine amidotransferase
64. stage V sporulation protein B
65. pseudouridylate synthase
66. amidohydrolase family protein
67. ATP-dependent DNA helicase PcrA, putative
68. uracil permease
69. drug resistance transporter, EmrB/QacA family
70. magnesium transporter
71. NADPH-dependent butanol dehydrogenase
72. oligopeptide transporter, OPT family
73. hypothetical protein (CPR_1285)
74. SsrA-binding protein
75. phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent
76. RNA polymerase sigma-54 factor
77. primosome assembly protein PriA
78. conserved hypothetical protein (CPR_1721)
79. L-asparaginase, type II
80. sensory box histidine kinase
81. transcriptional regulator, NrdR family
82. RNA polymerase sigma-G factor
83. cell division protein FtsZ
84. cell division protein FtsA
85. twitching motility protein PilT
86. peptidase, U32 family
87. conserved hypothetical protein (CPR_1743)
88. RNA-metabolizing metallo-beta-lactamase family protein
89. segregation and condensation protein B
90. segregation and condensation protein A
91. tyrosine recombinase XerD
92. NAD(+)/NADH kinase
93. hemolysin A
94. geranyltranstransferase
95. translation elongation factor P
96. pilus biogenesis protein, putative
97. diaminopimelate epimerase
98. methyltransferase, putative
99. stage V sporulation protein E
100. phospho-N-acetylmuramoyl-pentapeptide-transferase
101. UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase
102. udp-n-acetylmuramoylalanyl-d-glutamate--2,6-diaminopimelate
103. stage V sporulation protein D, spoVD, FtsI/pbp family
104. teichoic acid biosynthesis protein A
105. stage V sporulation protein D, spoVD, FtsI/pbp family
106. MutS2 family protein
107. potassium uptake protein, TrkH family
108. ribosomal protein L20
109. translation initiation factor IF-3
110. aspartate-semialdehyde dehydrogenase
111. dihydrodipicolinate reductase
112. cob(I)alamin adenosyltransferase, putative
113. 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase, putative
114. membrane proteins-like protein lmo0908, putative
115. bacterial extracellular solute-binding proteins, family 5 superfamily
116. bacterial type II secretion system protein F
117. secretion system protein E
118. hypothetical protein (CPR_2297)
119. ribosomal large subunit pseudouridine synthase f
120. 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
121. phosphoglucosamine mutase
122. sensor protein yycg
123. two-component response regulator
124. 4-alpha-glucanotransferase
125. transcriptional regulator, LacI family
126. conserved hypothetical protein (CPR_2346)
127. thioredoxin reductase
128. phosphoenolpyruvate-protein phosphotransferase
129. ribosomal protein S4
130. ribonucleotide-diphosphate reductase subunit beta

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 278-289 (2010)

131. sensor histidine kinase
132. DNA-binding response regulator
133. ribosomal protein S9
134. ribosomal protein L13
135. ABC transporter permease protein
136. ribosomal protein L17
137. DNA-directed RNA polymerase, alpha subunit
138. ribosomal protein S4
139. ribosomal protein S11
140. ribosomal protein S13p/S18e
141. preprotein translocase, SecY subunit
142. ribosomal protein L15
143. ATP-binding protein
144. hypothetical protein (CPR_0337)
145. DNA polymerase III, alpha subunit, interruption-C
146. cardiolipin synthetase
147. phosphopentomutase
148. quinolinate synthetase complex, A subunit
149. PTS system, IIB component
150. LacI family transcription regulator
151. PTS system, IIBC component
152. helicase/exonuclease
153. tRNA (guanine-N(7)-)-methyltransferase
154. ABC transporter domain protein
155. DNA-binding response regulator
156. glycosyltransferase, putative
157. beta-1,4-N-acetyl-mannosaminyltransferase, putative
158. polysaccharide transporter protein, putative
159. UTP-glucose-1-phosphate uridylyltransferase
160. UTP-glucose-1-phosphate uridylyltransferase
161. ABC transporter, permease protein
162. hypothetical protein (CPR_0491)
163. NAD-dependent 4-hydroxybutyrate dehydrogenase
164. permease
165. nitrite/sulfite reductase-like protein
166. stage V sporulation protein D
167. riboflavin biosynthesis protein RibD
168. riboflavin synthase, alpha subunit
169. riboflavin biosynthesis protein RibA
170. phosphate permease
171. fructose specific permease
172. 1-phosphofructokinase
173. diaminopimelate decarboxylase
174. amino acid ABC transporter, permease protein-like protein lin2352
175. cation efflux family protein
176. transporter-like protein lin1189
177. glycosyltransferase
178. polysaccharide biosynthesis protein, putative
179. DNA-binding response regulator
180. glutaminyl-tRNA synthetase
181. tyrosyl-tRNA synthetase
182. relA/spoT family protein
183. BirA bifunctional protein
184. degV family protein
185. rRNA methylase-like protein cspR
186. autolytic lysozyme, putative
187. sensory box histidine kinase
188. endonuclease III
189. serine O-acetyltransferase
190. amino acid (glutamine) ABC transporter, permease protein-like protein lin1851
191. amino acid ABC transporter ATP-binding protein
192. putative lipid kinase
193. phosphomethylpyrimidine kinase
194. transcriptional regulator, LacI family
195. d-galactose-binding periplasmic protein precursor
196. Galactoside transport ATP-binding protein mglA
197. beta-methylgalactoside transporter inner membrane component
198. fructose-1,6-bisphosphate aldolase, class II

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 278-289 (2010)

199. aminotransferase family protein
200. GTPase EngB
201. purine nucleoside phosphorylase
202. thiamine biosynthesis/tRNA modification protein ThiI
203. sodium/alanine symporter
204. cardiolipin synthase
205. glutamyl-tRNA reductase
206. oligoendopeptidase F
207. aminotransferase, class V
208. deoxyuridine 5'-triphosphate nucleotidohydrolase
209. sulfate permease, SulP family
210. nitrate ABC transporter ATP binding protein
211. uracil transporter
212. sucrose-6-phosphate hydrolase
213. sucrose operon repressor
214. spoVK domain protein
215. molybdopterin oxidoreductase
216. undecaprenyl diphosphate synthase, putative
217. ABC transporter (permease proteins)-like protein lmo1390
218. glycogen branching enzyme
219. DNA-binding response regulator
220. sensor histidine kinase
221. GTP pyrophosphokinase
222. protein-export membrane protein SecF
223. protein-export membrane protein SecD
224. holliday junction DNA helicase RuvB
225. holliday junction DNA helicase RuvA
226. conserved hypothetical protein (CPR_1923)
227. hypoxanthine phosphoribosyltransferase
228. membrane carboxypeptidase mrcB
229. sodium/alanine symporter family protein
230. spermidine/putrescine ABC transporter, permease protein (potB)-like protein
231. spermidine/putrescine ABC transporter, permease protein (potC)-like protein
232. spermidine/putrescine-binding periplasmic protein precursor
233. aspartate--ammonia ligase
234. lrgB-like family protein
235. aminopeptidase
236. hypothetical protein (CPR_1972)
237. hypothetical protein (CPR_1973)
238. RNA polymerase sigma factor RpoD
239. DNA primase
240. pyruvate,phosphate dikinase
241. GTP-binding protein Era
242. hypothetical protein (CPR_1998)
243. ribosomal protein L11 methyltransferase
244. heat-inducible transcription repressor HrcA
245. DNA polymerase III, delta subunit
246. ATP-dependent protease
247. deoxyribose-phosphate aldolase
248. manganese-dependent inorganic pyrophosphatase, putative
249. UDP-N-acetylglucosamine- -N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
250. DNA topoisomerase, GyrA/ParC subunit family
251. ROK family protein
252. sugar transport system (permease) (binding protein dependent transporter)
253. FemAB family protein
254. glutamine-binding periplasmic protein of glutamine ABC transporter
255. vncR, response regulator
256. ABC transporter (ATP binding protein)-like protein lmo1063
257. ribosomal protein S5
258. ribosomal protein L18
259. ribosomal protein L6
260. ribosomal protein S8
261. ribosomal protein L5
262. ribosomal protein L14
263. 50S ribosomal protein L16
264. 30S ribosomal protein S3
265. ribosomal protein L22
266. ribosomal protein L3

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 278-289 (2010)

267. ribosomal protein S10
268. ribosomal protein S7
269. ribosomal protein S12
270. ribosomal protein L7/L12
271. 50S ribosomal protein L10
272. ribosomal protein L1
273. ribosomal protein L11
274. transcription termination/antitermination factor NusG
275. thymidylate synthase, flavin-dependent
276. 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase
277. DNA repair protein RadA
278. transcriptional regulator, GntR family
279. thioredoxin-disulfide reductase
280. UDP-N-acetylmuramoylalanine--D-glutamate ligase
281. transcription elongation factor GreA
282. tRNA(Ile)-lysidine synthetase
283. MazG family protein
284. stage V sporulation protein B
285. thiamine biosynthesis protein ThiC
286. phosphoribosylaminoimidazole carboxylase, catalytic subunit
287. 3-dehydroquinate synthase
288. chorismate synthase
289. 3-dehydroquinate dehydratase, type II
290. hypothetical protein (CPR_0697)
291. probable permease, putative
292. Cytochrome C biogenesis protein transmembrane region family
293. ABC transporter, ATP-binding protein
294. cardiolipin synthetase
295. sodium:glactoside symporter family protein
296. antibiotic ABC transporter ATP binding protein SSO1934
297. acyl carrier protein phosphodiesterase
298. iron compound ABC transporter, permease protein
299. iron(III) dicitrate transport permease-like protein yusV
300. cymH protein
301. iron compound ABC transporter, permease protein
302. ferrichrome transport system, permease protein
303. Ferrichrome transport ATP-binding protein fhuC
304. ISCpe6, transposase orfA
305. NADH-dependent butanol dehydrogenase a
306. GntP family permease
307. uncharacterized conserved protein, YHAD family
308. MATE efflux family protein
309. thermophilic metalloprotease family protein
310. sucrose-6-phosphate hydrolase e1
311. PTS system, N-acetylglucosamine-specific IIBC component
312. cation efflux family protein
313. DNA-binding response regulator
314. pyridine nucleotide-disulphide oxidoreductase family protein
315. triple helix repeat-containing collagen
316. sensory box histidine kinase/response regulator
317. L-serine dehydratase, iron-sulfur-dependent, alpha subunit
318. oxidoreductase, 2-nitropropane dioxygenase family
319. HPr(Ser) kinase/phosphatase
320. methylglyoxal synthase
321. Predicted metal-dependent phosphoesterase (PHP family)
322. para-aminobenzoate synthase glutamine amidotransferase, component II
323. para-aminobenzoate synthase, component I
324. dihydropteroate synthase
325. dihydroneopterin aldolase/ 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase
326. thiazole biosynthesis protein ThiG
327. transporter, major facilitator family
328. methionine-R-sulfoxide reductase
329. transcriptional regulator, LacI family
330. teichoic acid linkage unit synthesis protein-like protein lin2663
331. ferripyochelin binding protein
332. HD superfamily hydrolase, YMDA
333. recA protein
334. CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 278-289 (2010)

335. DNA translocase FtsK
336. aspartokinase
337. riboflavin biosynthesis protein RibF
338. DHH subfamily 1 protein
339. DNA polymerase III, alpha subunit, Gram-positive type
340. 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase
341. 1-deoxy-D-xylulose 5-phosphate reductoisomerase
342. phosphatidate cytidylyltransferase
343. ribosome recycling factor
344. uridylate kinase
345. translation elongation factor Ts
346. ribosomal protein S2
347. GTP-sensing transcriptional pleiotropic repressor CodY
348. DNA protecting protein DprA
349. ribonuclease HII
350. ribosomal protein L19
351. radical SAM domain-containing protein
352. fatty acid/phospholipid synthesis protein PlsX
353. acetate kinase
354. phosphate acetyltransferase
355. pantetheine-phosphate adenylyltransferase
356. ATP-dependent DNA helicase RecG
357. DAK2 domain protein
358. GTPase YjeQ
359. protein phosphatase 2C family protein
360. polypeptide deformylase
361. pseudouridylate synthase family protein, yabo
362. nicotinate nucleotide adenylyltransferase
363. ribosomal protein L27
364. ribonuclease, Rne/Rng family
365. cell cycle protein, FtsW/RodA/SpoVE family
366. septum site-determining protein MinD
367. Penicillin-binding Protein dimerisation domain family
368. rod shape-determining protein MreB
369. DNA repair protein, RadC family
370. hypothetical protein (CPR_2114)
371. aminoacyl-histidine dipeptidase
372. hypothetical protein (CPR_2119)
373. PTS system, glucose-specific IIBC component
374. ABC transporter, substrate-binding protein
375. conserved hypothetical protein (CPR_2132)
376. glycoprotease family protein
377. preprotein translocase, SecA subunit
378. RecD/TraA family helicase
379. mbl protein
380. UDP-N-acetylglucosamine 1-carboxyvinyltransferase
381. UDP-N-acetylglucosamine 2-epimerase
382. uracil phosphoribosyltransferase
383. ribose 5-phosphate isomerase b
384. Sua5/YciO/YrdC/YwlC family protein
385. modification methylase, HemK family
386. thymidine kinase
387. transcription termination factor Rho
388. 4-diphosphocytidyl-2C-methyl-D-erythritol kinase
389. sensor histidine kinase HpkA
390. DNA-binding response regulator
391. aminoacyl-histidine dipeptidase
392. non-canonical purine NTP pyrophosphatase, rdgB/HAM1 family
393. ribonuclease PH
394. glucose-6-phosphate isomerase
395. oligopeptide ABC transporter, ATPase component
396. oligopeptide ABC transporter, ATPase component
397. oligopeptide ABC transporter, permease component
398. oligopeptide ABC transporter, permease component
399. transcription-repair coupling factor
400. PDZ domain protein
401. sensor histidine kinase
402. DNA-binding response regulator

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 278-289 (2010)

403.	UDP-N-acetylglucosamine pyrophosphorylase
404.	UDP-N-acetylmuramate--alanine ligase
405.	Transcriptional regulator
406.	metallopeptidase, family M24
407.	primase-like protein
408.	bifunctional acetaldehyde-CoA/alcohol dehydrogenase
409.	catabolite control protein A, putative
410.	glycerol uptake operon antiterminator
411.	glutamate racemase
412.	glutamine synthetase, putative
413.	Transcriptional regulator
414.	metallo-beta-lactamase family protein
415.	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
416.	YhiN family flavoprotein
417.	L-serine dehydratase, iron-sulfur-dependent, alpha subunit
418.	class II aldolase, tagatose bisphosphate family
419.	replicative DNA helicase
420.	ATP-dependent protease
421.	ribosomal protein L9
422.	DHH family protein
423.	stage 0 sporulation protein J
424.	sporulation initiation inhibitor protein soj
425.	methyltransferase GidB
426.	membrane protein OxaA

ISSN 0973-2063 (online) 0973-8894 (print)
Bioinformation 4(7): 278-289 (2010)