1. INTRODUCTION

1.1 *Clostridium perfringens:* Pathogenic micro-organism

Clostridium perfringens is a Gram-positive, spore-forming, heat-resistant, rod-shaped anaerobic bacterium, widely distributed in soil, marine sediment, decaying vegetation, and the intestinal tract of insects, humans and other vertebrates. *C. perfringens* is a member of the family Clostridiaceae of the Clostridiales, and is subdivided into five distinct serotypes (A to E), on the basis of each isolate's pattern of expression of four (alpha, beta, epsilon, and iota) of the 13 known *C. perfringens* toxins, which differ in their tropisms and toxigenicity. The virulence is derived largely from its prolific ability to express protein toxins and produce endospores. The spores produced by this organism persist in the environment and often contaminate raw food materials, causing a number of clinical conditions ranging from relatively mild food poisoning to the potentially life-threatening gas gangrene.

C. perfringens is in general linked to two types of food poisoning, namely, type A and type C. Type A is one of the most common human diseases caused by *C. perfringens*, which annually ranks among the leading foodborne diseases in industrialized countries. Type A induces a relatively minor gastrointestinal tract illness, causing diarrhoea and abdominal cramps for roughly 24 hours following an 8 to 12 hour incubation period. Type C food poisoning is a much more serious gastrointestinal tract disease, causing severe abdominal pain, dysentery (bloody diarrhoea), and vomiting, following a five to six hour incubation period. After these initial symptoms, the person may experience necrotic inflammation of the small intestine.

1.2 Novel Target Identification

Usually penicillin is used as one of the preferred antibiotics against the *Clostridium perfringens*, but it is only useful if the infection is diagnosed at early stages. However, comparison of in-vivo and in-vitro effectiveness of several antimicrobial agents including the penicillin's, indicate some discrepancies between these tests. The treatment of *C. perfringens* infections therefore may not be as straightforward as is commonly believed, and other antibiotics, or combinations thereof, may be more suitable than the penicillins. 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. So, the first step of developing new drug is to identify drug target. Here, accurate structural information of

validated target proteins provides a basis for the design and development of novel therapeutic agents. The increased number of high resolution protein structures available from the RCSB Protein Databank (PDB) has opened new opportunities for structure-based rational drug design. Still, the identification of potential protein binding pockets and occluded cavities remains a central issue, as the capability to interact with other proteins or small ligands determines the biological function of a protein. The size and shape of ligand binding sites and the distribution of functional groups in these pockets are of particular interest for the design of selective low-molecular weight ligands. This renders binding-site analysis pivotal for rational drug design, such as ligand docking or de novo molecular design. These methods require exact structural information of the binding-site as a starting-point. A variety of computational methods already exists for the location of possible ligand binding-sites. So, this study has been designed to assess the small molecule binding sites in pathogen *Clostridium perfringens*.

1.3 Pocket Scoring Algorithm

Once all the pockets have been identified, they are ranked according to Scoring Algorithm. The Algorithm uses the following six terms: 1) absolute residue sequence identity 2) sequence identity relative to the protein surface 3) absolute sequence identity with the closest human homologue 4) pocket volume 5) pocket surface area and 6) crystallographic resolution of the protein. This unique Pocket scoring approach incorporates accurate pocket identification along with sequence conservation with a similar organism, sequence conservation with the host and structure resolution. Sequence conservation between the *Clostridium perfringens* and similar organism was performed only on the pocket residues. Also the sequence conservation between each of the *C. perfringens* and their equivalent pockets in the most homologous human proteins was calculated to avoid the cross-reactivity with the host. As the algorithm utilizes, different criteria to identify potential binding site, we assume this provides the best analysis.

2. REVIEW OF LITERATURE

2.1 Pathogenic Bacteria *Clostridium perfringens*

Clostridium is a genus of Gram-positive bacteria, belonging to the Firmicutes. They are obligate anaerobes capable of producing endospores (Ryan, 2004). Individual cells are rod-shaped, which gives them their name. Clostridium consists of around 100 species that include common free-living bacteria as well as important pathogens (Wells, 1996). There are four main species responsible for disease in humans: *C. botulinum, C. difficile, C. tetani* and *C. perfringens.*

C. perfringens (**Fig 2.1**) formerly known as *C. welchii,* is an inhabitant of human normal intestinal flora, but still a pathogen responsible for many gastrointestinal illnesses with severity ranging from necrotizing enterocolitis in infants, food poisoning to gas gangrene (clostridial myonecrosis) in adults. And it is also responsible for many diseases in animals like enterotoxaemia in sheep and goats, lamb dysentery, ovine enterotoxaemia and pulpy kidney disease of sheep (Wells, 1996). *Clostridium perfringens* was first described in 1892 by Welch and Nutall as a Gram-positive, sporulating, obligate anaerobe and unusual among the pathogenic clostridia due to its non-motile nature. *Clostridium perfringens* is widely distributed in the environment and foods, and as stated above, forms part of the normal gut flora in both human and animals.

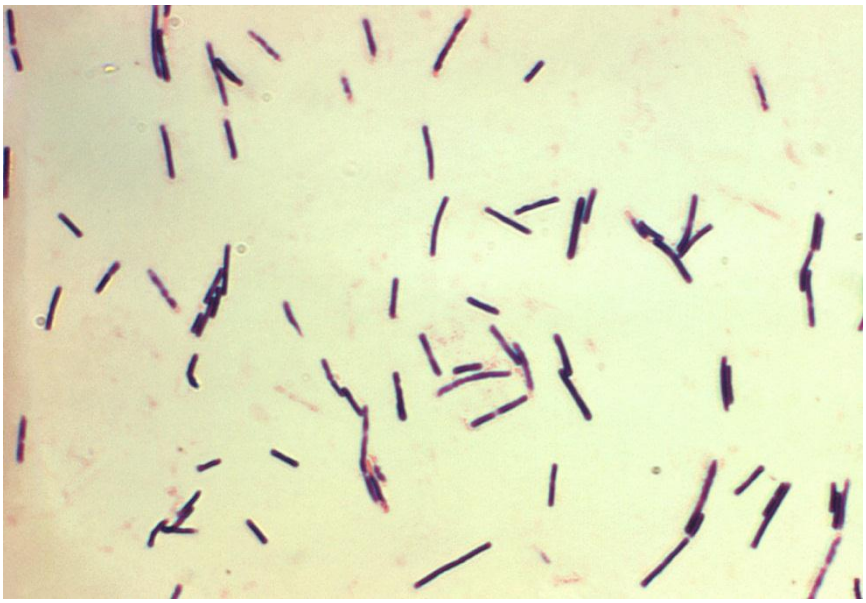


Fig 2.1 Photomicrograph of gram-positive *Clostridium perfringens* bacilli

Isolates of *C. perfringens* can be divided into five types (A to E) based on the particular extracellular toxins which they produce (Miyamoto, 2008) (**Table 2.1**). Each of these toxin types is responsible for specific disease syndromes (**Table 2.2**). *C. perfringens* diseases are generally mediated via the production of extracellular enzymes or toxins, with the exception of human food poisoning, which involves a sporulation-specific enterotoxin. The toxins implicated in gas gangrene, the most serious of these diseases, are a phospholipase C (α-toxin) and a thiol-activated hemolysin (θ-toxin or perfringolysin O). The role in pathogenesis of a variety of other extracellular enzymes such as collagenase (κ-toxin), hyaluronidase (μ-toxin), DNase (ν-toxin), and neuraminidase (sialidase) still remains to be precisely elucidated. The major toxin implicated in pulpy kidney disease is the potent ε-toxin, whereas lamb dysentery and human enteritis necroticans are primarily due to the effects of the β-toxin. The primary factor involved in *C. perfringens* food poisoning is an enterotoxin which traditionally has been thought of as a sporulation related protein (Rood, 1991).

Toxins → / ***C.perfringens* Type ↓**	**A**	**β**	**E**	**i**	**δ**	**θ**	**κ**	**λ**	**μ**	**N**	**Nm**	**En**
A	+	-	-	-	-	+	+	-	+	+	+	+
B	+	+	+	-	+	+	+	+	+	+	+	+
C	+	+	-	-	+	+	+	-	+	+	+	+
D	+	-	+	-	-	+	+	+	+	+	+	+
E	+	-	-	+	-	+	+	+	-	+	+	+

Table 2.1 Toxins produced by *C. perfringens* (Nm-neuraminidase or sialidase and En-Enterotoxin) (Rood, 1991)

C.perfringens	**Disease Produced**
A	Gas gangrene (Clostridial myonecrosis), Food poisoning, Necrotic entritis of infants
B	Lamb dysentry, Enterotoxemia of sheep, foals and goats
C	Human entritis necroticans (pigbel), Enterotoxemia of sheep, Necrotic entritis in animals
D	Enterotoxemia of sheep (Pulpy kidney disease)
E	Enteritis of rabbits

Table 2.2 Diseases caused by *C. perfringens* (Rood, 1991)

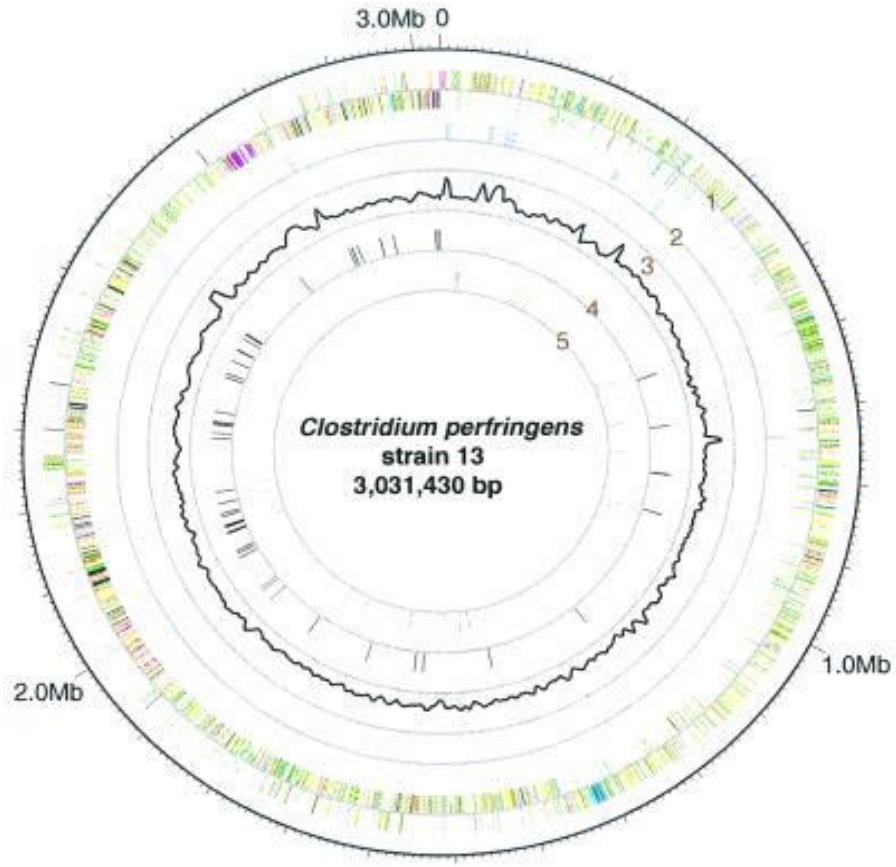


Fig 2.2 Circular map of the *C. perfringens* chromosome

The outer circle shows the scale. Ring 1 shows the coding sequence by strand (clockwise and counterclockwise); ring 2 shows the positions of rRNA genes; ring 3 shows G + C content (outer and inner rings correspond to 45% and 20%, respectively); ring 4 shows the positions of sporulation/germination genes; and ring 5 shows the positions of virulence-related genes. We used the following functional categories in ring 1: intermediary metabolism and fermentation (yellow), information pathways (pink), regulatory proteins (sky blue), conserved hypothetical proteins (orange), proteins of unknown function (light blue), insertion sequences and phage-related functions (blue), stable RNAs (purple), cell wall and cell processes (dark green), virulence (red), detoxification and adaptation (white), and sporulation and germination (black) (Shimizu, 2002)

C. perfringens is different from many other clostridia in that it is non-motile and, in vitro, forms spores only in specialized culture media. The organism is fermentative and grows rapidly in media containing carbohydrates. Under these conditions it produces copious amounts of H_2 and $C0_2$, which help to maintain an anaerobic environment. Owing to its rapid growth and relative aerotolerance, *C. perfringens* is easy to work with in the laboratory and has become a model organism for the development of clostridial genetics. In recent years there have been rapid advances in the genetics of *C. perfringens*. *C. perfringens* strain13, associated with human disease, has been shown to possess a single circular chromosome of about 3.58 Mb by PFGE (**Fig 2.2**). More than 100 restriction sites and 24 genetic loci have been located on the genome that is slightly larger in size than that of the other Gram-positive bacterium. It is highly significant that the genes for the α and θ toxins, which are produced by all serotypes of *C. perfringens*, are located near the presumed origin of replication, as this area should be among the most conserved regions of the genome.

2.2 Toxins and Virulence Factors

C. perfringens produces at least 16 potential virulence factors, including 12 toxins (α to ν), enterotoxin, hemolysin, neuraminidase (sialidase). Many of these virulence factors are simple hydrolytic enzymes secreted by the organism as part of its saprophytic life style in the soil, where they are probably involved in the putrefaction process. Their role in pathogenesis is therefore, at best, purely fortuitous **(Table 2.3)**.

1. α-Toxin or Phospholipase C

Numerous observations suggest that the most important virulence factor produced by clostridial myonecrosis (gas gangrene) causing isolates of *C. perfringens* is the α-toxin. Alpha-toxin possesses phospholipase C, sphingomyelinase biological activities that cause hemolysis, lethalitym, and dermonecrosis. The toxin is a zinc metallophospholipase that requires zinc for activation (**Fig 2.3**). First, the toxin binds to a binding site on the cell surface. The C-terminal domain of the protein toxin then binds a calcium ion, which allows the toxin to bind to the phospholipid head-groups on the cell surface. The C-terminal domain enters the phospholipid bilayer. The N-terminal domain has phospholipase activity. This property allows hydrolysis of phospholipids such as phosphatidyl choline, thus mimicking endogenous phospholipase C of the signal transduction pathway. The hydrolysis of phosphatidyl choline produces diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP_3) which activate a variety of second messenger pathways.

TOXIN	GENE LOCATION	BIOLOGICAL ACTIVITY
Alpha	Chromosome	Cytolytic, Hemolytic, Dermonecrotic, Lethal
Beta	Plasmid	Cytolytic, Dermonecrotic, Lethal
Beta 2	Plasmid	Cytotoxic, Lethal
Epsilon	Plasmid	Edematous in Liver, Kidney, Nervous system
Iota (Ia & Ib)	Plasmid	Disruption of actin cytoskeleton and affects cell membrane integrity
Theta	Chromosome	Hemolytic, Modulation of Host Inflammatory response
Enterotoxin	Chromosome/Plasmid	Cytolytic, Lethal, Leakage of water and ions

Table 2.3 Toxins of *Clostridium perfringens* (Perelle, 1993)

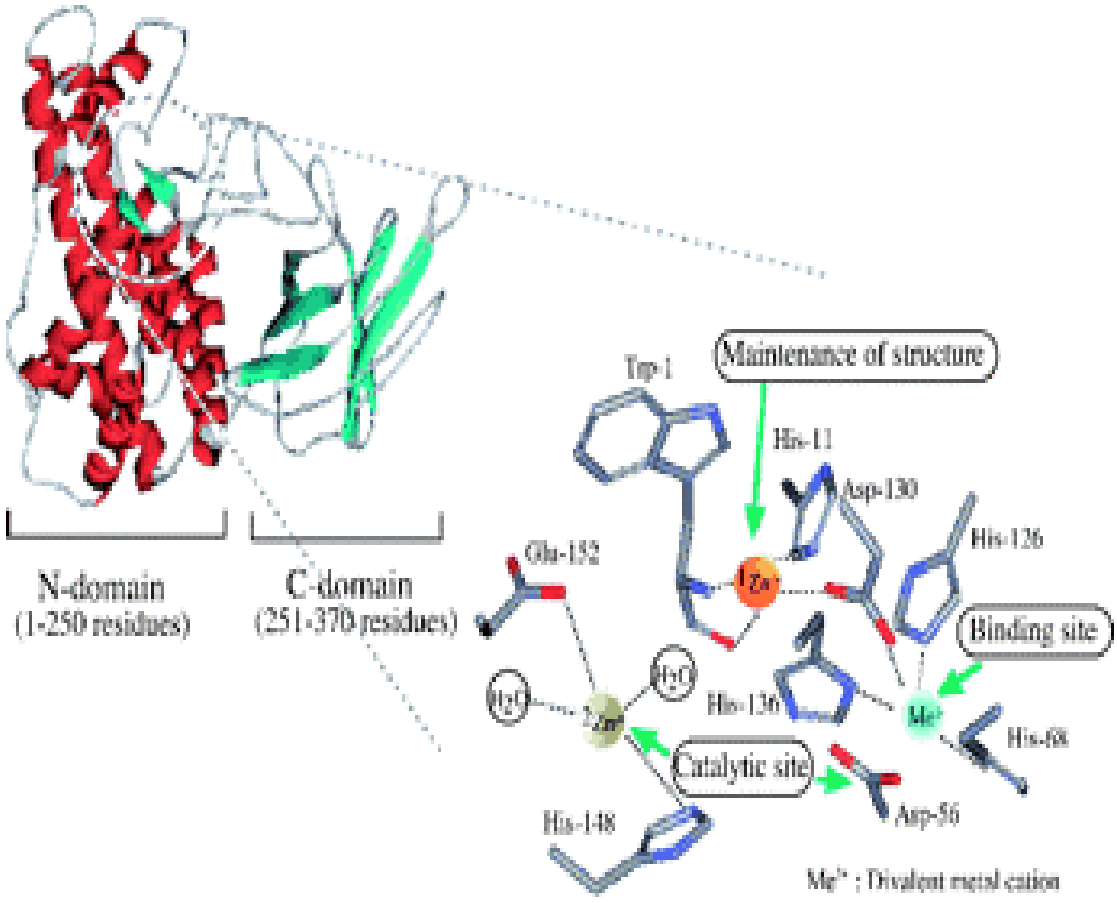


Fig 2.3 Protein model of alpha toxin showing the two domains (Sakurai, 2004)

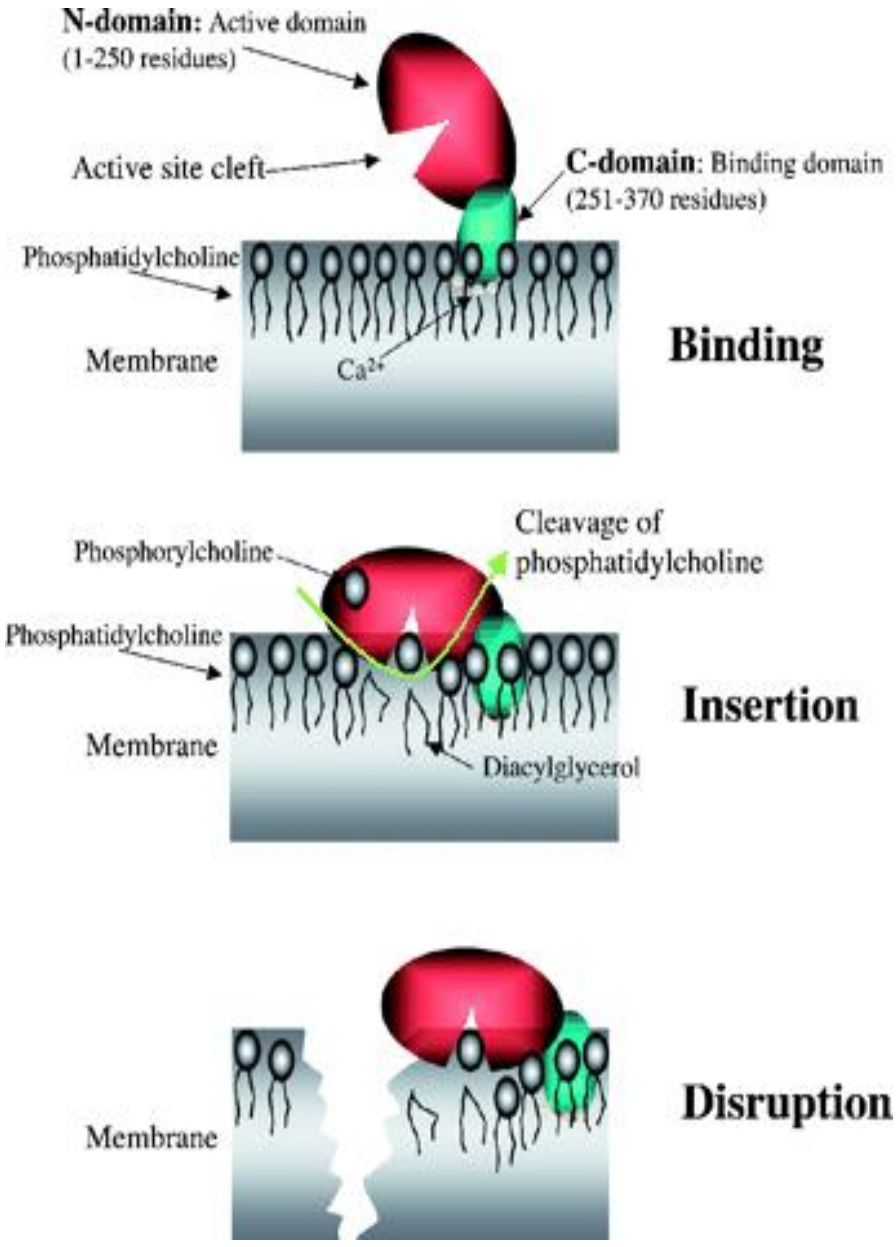


Fig 2.4 Mode of membrane destruction by alpha toxin. (Sakurai, 2004)

The end result includes activation of arachidonic acid pathway and production of thromboxane A2, interleukin-8 (IL-8; involved in the inflammatory response, causing host damage), platelet-activating factor, and several intercellular adhesion molecules. These actions combine to cause edema due to increased vascular permeability. *C. perfringens* strains that produce various gelatinolytic enzymes such as collagenase, a protease that cleaves native collagen in its triple helical conformation, also play a role in the physical attributes associated with gas gangrene (Matsushita, 1994). The structural breakdown of collagen - the framework of muscle protein - facilitates the formation gas gangrene. Consequently, the collagenase produced by this pathogen is involved in tissue necrosis, along with other toxins such as phospholipase C and thiol-activated haemolysin. Once the targeted cells have died due to infection, *C. perfringens* bacteria scavenge dead human cells for carbon, producing gas as a byproduct of metabolism. In fact, *C. perfringens* bacteria produce a relatively large quantity of carbon dioxide, causing further tissue damage due to physical pressure (**Fig 2.4**).

2. β-Toxin

C. perfringens types B and C produce β-toxin, which is both lethal and necrotic (**Table 2.1**). Although type B strains cause concern in veterinary medicine, type C isolates can be important in human disease and are responsible for necrotic enteritis, which has decimated poorly nourished individuals in postwar Germany and New Guinea, where the disease is known as darmbrand and pigbel, respectively. The β-toxin has been purified to homogeneity and shown to correspond to a heat-sensitive 28-kDa protein which is highly sensitive to trypsin. This observation is of physiological significance since in normal individuals the toxin is rapidly inactivated by trypsin in the small intestine. Clearly, in times of starvation, trypsin levels are low and the levels of active β-toxin in the gut are therefore much higher and it can lead to severe type of food-poisoning. Generally, human endothelial cells are sensitive to the toxic effect of beta-toxin, since it forms potential-dependent, cation-selective channels in lipid bilayers, inducing the release of arachidonic acid and leakage of inositol from these cells, in addition to the efflux of potassium ions and the influx of calcium, sodium, and chlorine ions.

3. ε-Toxin

Like the α, β, and i-toxins, the ε-toxin is both lethal and necrotizing, although its precise biological activity has not yet been identified. It is produced by type B and D strains (Miyamoto, 2008) and is of great veterinary interest because it causes a rapidly fatal enterotoxemia amongst ungulates which is commonly referred to as pulpy kidney or overeating disease. The toxin is produced in precursor form and is activated following the

proteolytic removal of a 14-aminoacid residue peptide from the NH2 terminus of the protoxin. Proteolysis can result from the action of *C. perfringens* proteases, such as the λ-toxin, or the effect of trypsin in the gut. Among the symptoms provoked by ε-toxin are increased intestinal permeability, lung edema, and excess pericardial fluid. Its most striking effect is on the kidneys, which become swollen, pulpy a few hours before death. As immunity can be conferred by vaccination with a toxoid preparation, the gene for ε-toxin, represented a major target for biotechnologists. This gene has recently been cloned and sequenced to produce a second-generation veterinary vaccine.

4. i-Toxin

The i-toxin is the last of the four major lethal toxins produced by *C. perfringens* and is normally associated with type E strains (Perelle, 1993)(**Table 2.1**). It provokes necrosis and, like ε-toxin, is reportedly produced from a protoxin by proteolysis. The toxin was purified and found to comprise two distinct polypeptide chains which acted synergistically in mouse lethality and dermonecrosis assays. Subsequent studies showed the smaller polypeptide to possess ADP-ribosylating activity (Damme, 1996), and actin was one of the substrates modified.

5. θ-Toxin

All five types of *C. perfringens* produce a lethal hemolysin, θ-toxin, which is also known as θ-hemolysin, perfringolysin O, or the thiol-activated cytolysin. The primary sequence of θ-toxin, deduced from the pfoA gene sequence, revealed a 494-residue preprotein from which a 28-residue signal peptide is removed on secretion to the medium. Mature θ-toxin has a predicted molecular weight of 52,469 Da and, when purified from *C. perfringens* or *E. coli*, appears as a 54,000-Da species. θ-toxin is a member of a family of thiol-activated cytolysins produced by a diverse group of gram-positive bacteria including the Bacillus, Streptococcus, Clostridium, and Listeria genera (**Fig 2.5**). These cytolysins use cholesterol as a receptor (which play multiple roles: targeting, promotion of oligomerization, triggering a membrane insertion competent form, and stabilizing the membrane pore) and form large pores in target cell membranes (Shepard, 2000). They are reversibly inactivated by oxidation. To date, the primary structures of five of these cytolysins (alveolysin, listeriolysin, streptolysin, pneumolysin, and perfringolysin) have been deduced from a cloning and sequencing approach. The θ-toxin plays a role in the tissue necrosis associated with *C. perfringens* gas gangrene and is responsible for the depletion of polymorphonuclear leukocytes in the affected zone. Low concentrations of θ-toxin cause altered polymorphonuclear leukocyte morphology, metabolism, and migration.

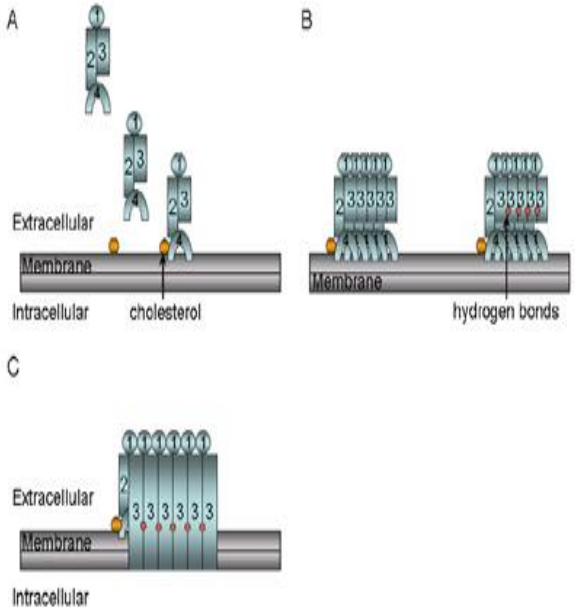


FIG 3.5 The mechanism of cholesterol-dependent cytolysin pore formation
Cholesterol-dependent cytolysins bind cholesterol containing membranes and lyse these cells, presumably by forming pores in the plasma membrane. These proteins have four domains; domain 3 and 4 are the only domains that interact with the membrane. First, soluble monomers attach to the cell membrane in proximity to cholesterol and oligomerize on its surface (**A**). A ring or arch-like structure is created by the oligomerization of the monomers on the surface of the membrane. Hydrogen bonds form between the domain 3 β-hairpins of the different monomers (**B**). The oligomers form a pore by simultaneously inserting into the membrane (**C**). (Shepard, 2000).

6. Enterotoxin

Of the major virulence factors produced by *C. perfringens*, the enterotoxin is one of the best studied. *C. perfringens* is the etiological agent of two very different types of food poisoning affecting humans and is the third most important source of food-borne disease after *S. aureus* and Salmonella species (Czeczulin, 1996). Some type A strains are associated with a mild form of food poisoning, resulting from enterotoxin production, whereas type C strains are responsible for necrotic enteritis, a rare but often deadly intestinal disease provoked by production of the β-toxin. *C. perfringens* food poisoning occurs following ingestion of heavily contaminated food, usually meat products containing more than 106 CFU/g. *C. perfringens* cells sporulate in the small intestine and concomitantly generate large quantities of enterotoxin. This can result, 12 to 24 hour later, in intense nausea, diarrhoea, and, more rarely, vomiting, with the occasional fatality amongst the elderly and debilitated. Infected individuals generally recover rapidly and are unaware of the origin of their malaise. In addition, it has also been suggested that *C. perfringens* enterotoxin may be involved in antibiotic-associated diarrhoea.

The enterotoxin accumulates in large quantities intracellularly, where, owing to its limited solubility, it can form inclusion bodies. Following lysis of the bacteria in the gut, the toxin is liberated and, like the ε-toxin, processed to a more active form. The current structural model for enterotoxin suggests a two-domain organization with the N-terminal region required for cytotoxicity while the C terminus interacts with cell membrane receptors, one of which is a 50-kDa protein species (Gao, 2011). It is clear that our understanding of enterotoxin function and molecular organization will benefit from further structural studies.

2.3 Antibiotic Resistance Determinants

Penicillin has traditionally been regarded as the drug of choice for the treatment of *C. perfringens* gas gangrene. However, doubts have been cast upon the efficacy of penicillins and cephalosporins in the clinical situation. Comparison of in vivo and in vitro effectiveness of several antimicrobial agents including the penicillins, indicate some discrepancies between these tests. The treatment of *C. perfringens* infections therefore may not be as straightforward as is commonly believed, and other antibiotics, or combinations thereof, may be more suitable than the penicillins.

Recently attention has been focused on the role of widespread use of antimicrobials in growth promotion and therapy of infections in food-producing animals as a potential transfer route of antimicrobial-resistant bacteria or the genes encoding antimicrobial resistance into

the human food chain (Piddock, 2000). As a consequence, resistance is most common where there is heavy use of antimicrobials and appreciable host-to-host contact. Humans and animals live in close association with large numbers of bacteria, most of which are found in the large intestine, where they are exposed to antimicrobials, exchange genetic material with other bacteria and, on excretion, contaminate the environment or colonize other animals and humans. Since *C. perfringens* is an indicator bacterium that constitutes a natural part of the intestinal flora of both humans and animals, using this bacterial species makes it feasible to compare the levels of resistance between populations. So, overall, *C. perfringens* showed the highest resistance to tetracycline (56.2%), followed by imipenem (24.9%), metronidazole (9.5%), penicillin G (9%), vancomycin (4.5%), chloramphenicol (3%) and ceftriaxone (1%). The majority of the isolated strains from pig feces (77.8%), environment (72.7%), human feces (44.9%) and food (28%) showed resistance to tetracycline. Strains isolated from human feces only showed low resistance to ceftriaxone (2.5%) and vancomycin (10.1%). While, Penicillin G had high activity and low rate of resistance (10-12% for strains isolated from humans, animals and food). Among 62.7% of antimicrobial resistant strains, 39.3% were resistant to a single drug and 23.4% were multiple-drug resistant (MDR) (Tansuphasiri, 2005).

Analysis of the published literature had showed that the main determinants of antibiotic resistance are irrational use of antibiotic drugs in humans and animal species, insufficient patient education when antibiotics are prescribed, lack of guidelines for treatment and control of infection in hospitals, lack of awareness and relevant scientific information for physicians prescribing antibiotics, and lack of official government policy on the rational use of antibiotics in the public and private sectors. It is therefore necessary that every country take proper measures to control these determinants in order to check antibiotic resistance (Franco, 2009).

2.4 Identification of new Drug Targets

Bacterial diseases are a major contributor to global morbidity and mortality. The emergence of bacteria resistant to frontline antibiotics and the ability of bacteria to acquire and spread resistance are drivers for the development of novel antibiotics. Antibiotic discovery strategies have adapted to utilise the wealth of genetic sequence data recently available in publicly accessible databases. These databases provide substantial amount of data that requires the development of in-silico techniques for the prediction, identification and selection of targets for investigation (Moellering, 2011). Current antibiotics have often

targeted proteins involved in cell maintenance or structure. However, more recently, essential genes and their protein products have become of particular interest, essentiality being one of the characteristics used for identification and prioritisation drug target (**Table 2.4**).

Target Feature	Benefit as drug target
Essential for the survival of the organism	Disruption of the gene or the gene product results in bacterial death
Conserved in a number of different Bacteria	Selects targets that are likely to be broad spectrum rather than species specific
Involvement in crucial pathway	The pathway can be interrupted
Druggable	Interacts and binds with a small molecule or ligand indicating there is a binding site that can be blocked
Low molecular weight and without transmembrane domains	Smaller proteins are more likely to be soluble and easier to purify, membrane proteins are specially difficult to purify
Characterized protein	Including functional and structural characterization further benefits molecule selection
No homology to human or other mammalian proteins	Since an eventual drug will be used in humans, a target specifically bacterial is preferred to avoid toxicity or detrimental interactions with human proteins resulting in side effects

Table 2.4 Selection criteria for new antibacterial targets (Cooper, 2011)

Essential genes are by definition necessary for replication and viability, and therefore the deletion, interruption or blocking of the protein expressed by an essential gene, results in death of the organism, making them attractive targets for drug development. By identifying a protein essential to a bacteria and developing inhibitors to that protein, new antimicrobials can be identified. Essential genes that are conserved across bacterial genera have been proposed as promising candidates for broad spectrum drug targets, active against multiple bacterial species.

There are wide range of bacterial toxins, which have been described, ranging from peptides to complex high molecular weight proteins and lipopolysaccharides. The vast majority of toxins that play a significant role in the pathogenesis of disease are proteins, and often they are enzymes. In some diseases these toxins play major roles in the pathogenesis of disease, and mutants which are unable to produce these toxins are attenuated. In case of *Clostridium perfringens*, the alpha-toxin produced by them plays a key role in the

development of gas gangrene by modulating the host inflammatory response and reducing the blood supply to infected tissues. Mutants of *C. perfringens* which are unable to produce the toxin are unable to cause disease. To gain the understanding of the mode of action of toxins, as well as of other essential proteins, we have Crystallography as a new valuable tool, which provide deep knowledge of molecular basis of action of proteins. In turn this facilitates further research leading to the development of compounds that are able to block toxicity. Therefore, the development of computational methods to model protein–protein docking, identifying promising binding pockets, and predicting protein-ligand association facilitate the discovery of small molecules capable of inhibiting pathogen, a major new challenge in drug design. Still, identification of novel drug-binding targets is an insufficiently met need in drug discovery. Accurate structural information of validated target proteins provides a basis for the design and development of novel therapeutic agents. The increased number of high resolution protein structures available from the RCSB Protein Databank (PDB) has opened new opportunities for structure-based rational drug design. Still, the identification of potential protein binding pockets and occluded cavities remains a central issue, as the capability to interact with other proteins or small ligands determines the biological function of a protein. The size and shape of ligand binding sites and the distribution of functional groups in these pockets are of particular interest for the design of selective low-molecular weight ligands. This renders binding-site analysis pivotal for rational drug design, such as ligand docking or de novo molecular design. These methods require exact structural information of the binding site as a starting point. A variety of computational methods already exists for the location of possible ligand binding-sites. Most of these pocket detection algorithms solely rely on geometric criteria to find clefts and surface depressions (Weisel, 2007). Experimental and computational prediction of binding pockets on the surface of proteins has been successfully applied to rational drug design, and thus they can be one of the first choice computational tools to characterize a protein–protein interface in search of potential pockets (Grosdidier, 2009).

Traditionally, ligand binding sites have been elucidated through co-crystallization with a known ligand or through site-directed mutagenesis of putative active site residues. However, if such data is not available, then other means must be explored. An alternative method of determining binding sites is through computational structural analysis. This type of approach has been shown recently to detect successfully greater than 95% of the ligand binding sites in a benchmark of over 5000 protein structures. Once potential ligand binding sites are identified, the next question is whether small molecules bound to those sites will

interfere with the function of the protein. One method for determining the functional significance of a protein region is through sequence conservation to related homologues. This method has been shown to be particularly useful in determining interaction sites on proteins. Due to the unfulfilled requirement of target discovery, we decided to approach this problem by surveying all proteins in a genome-wide scale. Here we present a methodology for the comprehensive computational evaluation of all small-molecule binding sites across a genome, and the potential to target these sites for structure-based drug design.

3. MATERIALS AND METHODS

The whole experimental procedure based on in-silico analysis and it involves various web based tools, Desktop Application Software with the use of database as a data tier application and is compatible on both the Platform such as Windows 7 Professional, 4GB RAM, Service Pack 2 and Red Hat Linux Client 5.0.

3.1 UniProtKB/Swiss-Prot and TrEMBL

UniProt is the Universal Protein resource, a central repository of protein data created by combining the Swiss-Prot, TrEMBL and PIR-PSD databases. It is a comprehensive, high-quality and freely accessible database (http://www.uniprot.org) which contains a large amount of information about the biological function of proteins derived directly from the research literature.

UniProt Consortium: The UniProt Consortium comprises the European Bioinformatics Institute (EBI), the Swiss Institute of Bioinformatics (SIB), and the Protein Information Resource (PIR). EBI located at the Wellcome Trust Genome Campus in Hinxton, UK, hosts a large resource of bioinformatics databases and services. SIB located in Geneva, Switzerland, maintains the ExPASy (Expert Protein Analysis System) servers that are a central resource for proteomics tools and databases. PIR hosted by the National Biomedical Research Foundation (NBRF) at the Georgetown University Medical Center in Washington, DC, USA, is heir to the oldest protein sequence database, Margaret Dayhoff's Atlas of Protein Sequence and Structure, first published in 1965. In 2002, EBI, SIB, and PIR joined forces as the UniProt Consortium. Each consortium member is heavily involved in protein database maintenance and annotation.

UniProt Knowledgebase (UniProtKB) is a protein database curated by experts, consisting of two sections: UniProtKB/Swiss-Prot (containing reviewed, manually annotated entries) and UniProtKB/TrEMBL (containing unreviewed, automatically annotated entries).

1. **UniProtKB/Swiss-Prot**

UniProtKB/Swiss-Prot is a high-quality, manually annotated, non-redundant protein sequence database. It combines information extracted from scientific literature and biocurator-evaluated computational analysis. The aim of UniProtKB/Swiss-Prot is to provide all known relevant information about a particular protein. Annotation is regularly reviewed to keep up with current scientific findings. The manual annotation of an entry involves detailed analysis of the protein sequence and of the scientific literature (Apweiler, et al., 2004).

2. UniProtKB/TrEMBL

UniProtKB/TrEMBL contains high-quality computationally analyzed records, which are enriched with automatic annotation. It was introduced in response to increased dataflow resulting from genome projects, as the time- and labour-consuming manual annotation process of UniProtKB/Swiss-Prot could not be broadened to include all available protein sequences. The translations of annotated coding sequences in the EMBL-Bank/GenBank/DDBJ nucleotide sequence database are automatically processed and entered in UniProtKB/TrEMBL. UniProtKB/TrEMBL also contains sequences from PDB, and from gene prediction, including Ensembl, RefSeq and CCDS.

For *Clostridium perfringens*, in total 38 protein sequence with complete structural information are available at UniProt. Amongst these 38 structural proteins, 15 were selected for further studies on the basis of their relevance and importance in bacteria's life cycle (**Table 3.1**). Some of the selected proteins are important for pathogenesis (Toxins/Enzymes) and other are involved in metabolic processes like nucleotide biosynthesis, isoprenoid biosynthesis, nucleotide binding and other catalytic and transport activities.

3.2 Protein Data Bank (PDB)

The Protein Data Bank (PDB) is a repository for the 3-D structural data of large biological molecules. The data is typically obtained by X-ray crystallography or NMR spectroscopy and submitted by biologists and biochemists from around the world can be accessed at no charge on the internet at web-address (http://www.rcsb.org/pdb/).

The PDB originated in 1971, by Walter Hamilton of the Brookhaven National Laboratory who agreed to set up the data bank at Brookhaven. Upon Hamilton's death in 1973, Tom Koeztle took over direction of the PDB. In January 1994, Joel Sussman was appointed head of the PDB. In October 1998, the PDB was transferred to the Research Collaboratory for Structural Bioinformatics (RCSB). Helen M. Berman of Rutgers University was appointed as new director. In 2003, with the formation of the wwPDB, the PDB became an international organization (Bernstein, 1977).

The Protein data bank stores all the information regarding protein molecule and each protein information can be accessed by unique identification ID. The data available can be obtained in pdb file format or xml file format and in order to view the structure files one of the several open source computer program like Rasmol, Pymol, Webmol, Jmol etc. could be used, some of which are freely available.

Entry	Protein names	Cross-reference (PDB)	Gene names
Q8XLE8	Phosphoenolpyruvate carboxylase	**3ODM**	ppcA
P54965	Choloylglycine hydrolase (CBAH)	2BJF, **2BJG**, 2RF8, 2RG2, 2RLC	cbh
Q0TUN8	Energy-coupling factor transporter ATP-binding protein	**3GFO**	cbiO1
P01558	Heat-labile enterotoxin B chain	2QUO, **2XH6**, 2YHJ, 3AM2	Cpe
P26831	Hyaluronoglucosaminidase (Mu toxin)	**2OZN**, 2W1Q, 2W1S, 2W1U, 2WDB	nagH
P0C216	Phospholipase C (PLC)(Alpha-toxin)	**1CA1**, 1GYG, 1QM6, 1QMD, 1SB4	Plc
Q0TTB4	Phosphoribosylaminoimidazole-succinocarboxamide synthase (SAICAR synthetase)	**3NUA**	purC
P0C2E9	Perfringolysin O (Theta-toxin)	1M3I, 1M3J, **1PFO**, 2BK1, 2BK2	pfoA
Q57398	Epsilaon-toxin (Epsilon toxin)	**1UYJ**	Etx
Q46220	Iota toxin component Ia	1GIQ, **1GIR**, 3BUZ	
Q8XIQ9	Probable manganase-dependent inorganic pyrophosphatase	3L2B, **3L31**	CPE2055
Q8XJE0	Geranyltranstransferase	**3UCA**	CPE1820
Q8XI95	Putative uncharacterized protein	**3O6U**	CPE2226
Q0TRG9	Malonyl CoA-acyl carrier protein transacylase	**3PTW**	fabD
Q0TT74	DHHA1 domain protein	**3KEW**	CPF_0714

Table 3.1 Proteins with structural knowledge selected from UniProt KB and short-listed from Protein Data Bank Information (Highlighted PDB ID)

The PDB file format is a textual file format which describes the three dimensional structures of molecules held in the Protein Data Bank. Most of the information in database pertains to proteins, and the pdb format accordingly provides for rich description and annotation of protein properties. It usually contains the following records-

- ATOM records: describe the Atom name, residue name and then coordinates of the atom i.e.; x, y and z coordinates and units in Angstroms. The next three columns are the occupancy, temperature factor, and the element name, respectively.
- HEATM records: describe coordinates of hetero-atoms, that is those atoms which are not part of the protein molecule.
- SEQRES records: describes the sequences of the peptide chains (named A, B and C)
- REMARK records: can contain free-form annotation, but they also accommodate standardized information
- HEADER, TITLE and AUTHOR records: provide information about the researchers who defined the structure; numerous other types of records are available to provide other types of information.

Molecular models based on crystallographic PDB entries with released coordinates were used. In cases where multiple entries existed for a given protein, those with highest resolution, sequence identity closest to wild-type, and bound ligands were chosen with higher priority. For PDB entries representing more than one instance of the biologically functional quaternary structure, the molecule with the best overall similarity to the other molecules was used. The missing subunits need to be generated using crystallographic symmetry transformations and quaternary structure information from the literature (**Table 3.1**).

3.3 MolProbity

MolProbity is implemented in PHP as a web server located at http://molprobity.biochem.duke.edu. It provides a graphical interface to a collection of Richardson lab programs for validation and structure correction.

The atomic models of proteins and nucleic acids that come from X-ray crystallography and NMR are our most accurate sources of 3D information about these molecules, far more reliable than computed structures obtained from modeling or simulation. They are best when determined at high resolution or with many restraints per residue, but even then are not perfect, nearly all structures in the Protein Data Bank have a few local errors, such as backwards-fit branched side chains, flipped amides and imidazoles, incorrect sugar puckers, misoriented ligands, misidentified ‘waters’ and local errors in chain tracing.

Such errors are usually due to misinterpretations of ambiguous experimental data. The ambiguity can often be resolved by considering additional information (such as steric interactions between atoms). This is the kind of structure validation data provided by MolProbity (Randy, 2011).

MolProbity is related to programs such as PROCHECK, PROCHECK-NMR, WHATIF and OOPS, which provide both overall statistical evaluations and flags of local problem areas, concentrating primarily on geometrical measures that can be analyzed from the model. Whereas global validation measures serve the function of judging whether a structure meets accepted current practice, local measures are especially important to users of structures, since no level of global quality can guarantee protection against a large local error in the region of specific interest. MolProbity is unique in offering all-atom contact analysis and up-to-date, high-accuracy Ramachandran and rotamer distributions. It is also broader in scope than many validation programs, as it applies to both X-ray and NMR structures, and to both proteins and nucleic acids. Finally, it is useful to both 'consumers' and 'producers' of structural models: consumers can check that regions of interest are accurate, and producers can find and fix errors during fitting and refinement (Vincent, 2010).

The primary input for MolProbity is structural models in PDB format. Models may be uploaded directly, or MolProbity can fetch them from the PDB, on giving the appropriate identifier. Models may come from crystallography, NMR or computation. MolProbity then uses a variety of physics and knowledge based algorithms to analyze a structure. The primary basis of its enhanced effectiveness is all-atom contact analysis, as implemented in Probe (**Fig 3.1**). All-atom contacts are exquisitely sensitive to a wide variety of local misfittings, but they are not yet available in other validation systems. They do require explicit hydrogen atoms, but MolProbity can add and optimize these using Reduce, while at the same time detecting and automatically fixing flipped Asn, Gln and His sidechains (**Fig 3.2**). MolProbity also uses carefully filtered, high-accuracy Ramachandran and rotamer distributions to check mainchain and sidechains for conformational outliers. The different types of analysis are synthesized into two integrated reports on the structural model: one tabular and one graphical. MolProbity illustrates the proposed corrections with 3D kinemage graphics and allows the user to veto changes before proceeding. Using MolProbity server, Structure validation of selected PDB structures was done.

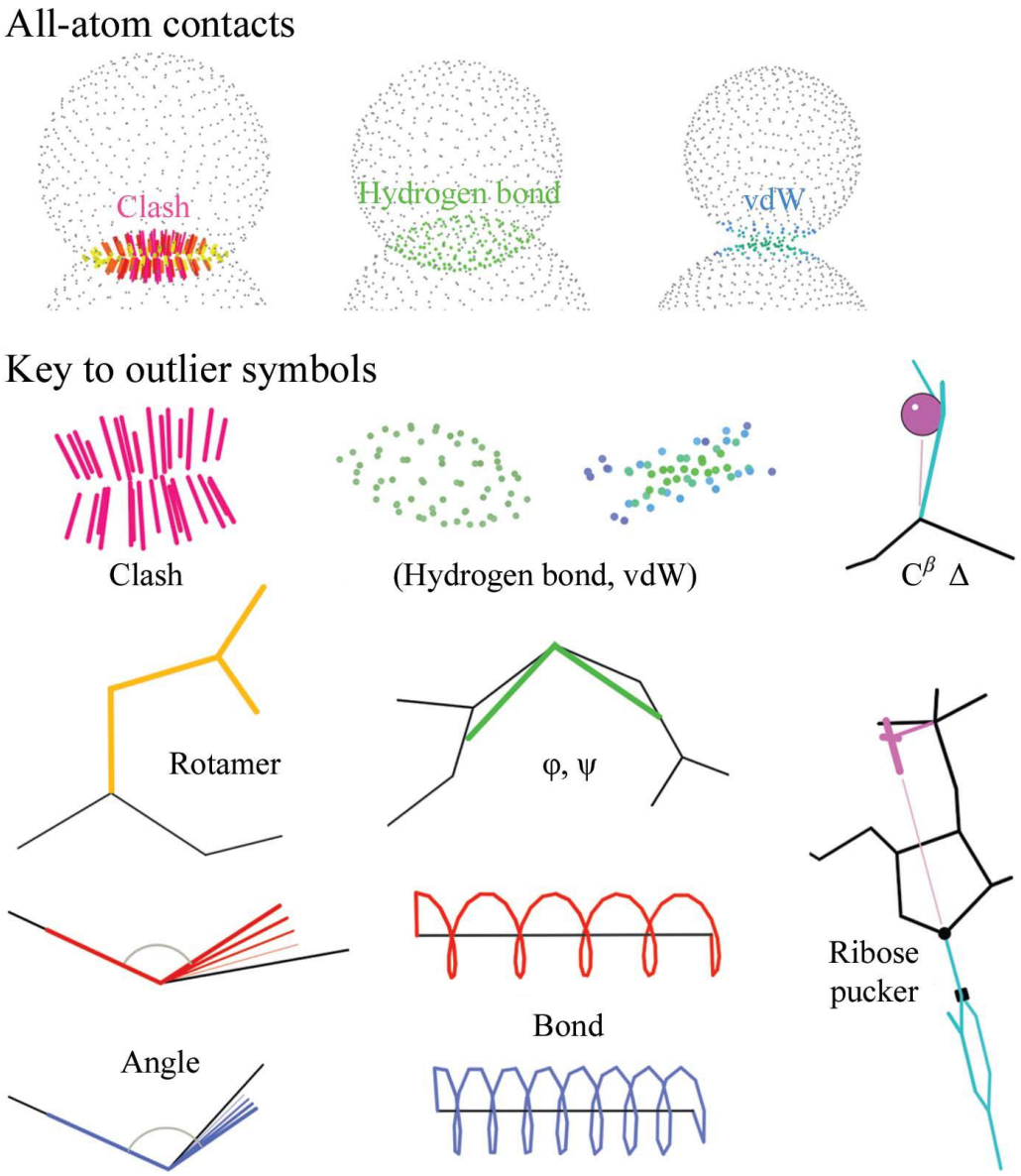


Fig 3.1 Symbols used in MolProbity (Vincent, 2010)

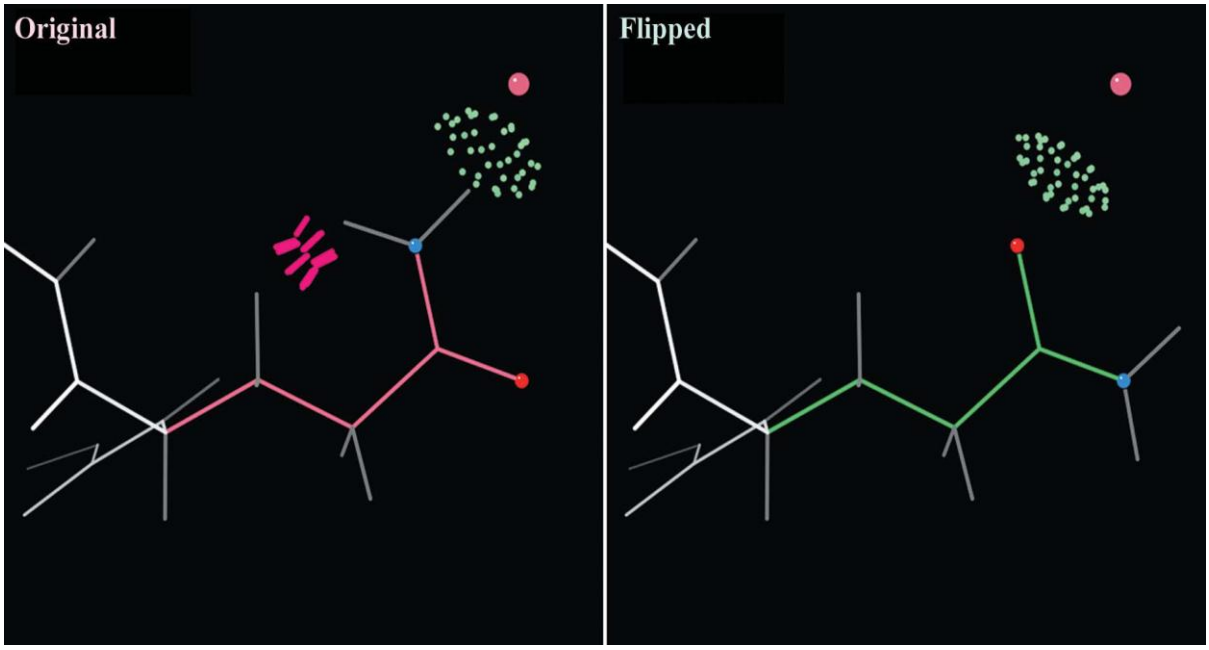


Fig 3.2 The simple 'flip' correction of a Gln side-chain amide in protein structure (Vincent, 2010)

3.4 PyMOL

PyMOL is an open-source, user-sponsored, molecular visualization system created by Warren Lyford DeLano and commercialized by DeLano Scientific LLC, which is a private software company dedicated to creating useful tools that become universally accessible to scientific and educational communities (http://www.PyMol.org). So, it is one of the visualization tools available for use in structural biology. It is well suited to producing high quality 3D images of small molecules and biological macromolecules such as proteins. The Pymol portion of the software's name refers to the fact that it extends, and is extensible by the Python programming language (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

PyMOL was originally designed to:

- Visualize multiple conformations of a single structure [trajectories or docked ligand ensembles]
- Interface with external programs
- Provide professional strength graphics under both Windows and Unix
- Prepare publication quality images, and
- Fit into a tight budget.

All of these goals have since been realized. PyMol was used to prepare all the publication quality images of protein structures.

3.5 Computed Atlas of Surface Topography of Proteins (CASTp)

Proteins in order to perform their function interact with other molecules, including other proteins, ligands, substrates and DNA. The three dimensional shape of a protein is a key factor in these interactions, providing docking sites or blocking access to particular amino acids. So, knowledge of protein surface structure enables detailed studies of the relationship of protein structure and function. Specifically, characterization of protein surface regions helps to analyze enzyme mechanism, to determine binding specificity and to plan mutation studies. It can also facilitate the identification of the biological roles for newly solved protein structures with an unknown function (Dundas, 2006).

CASTp is a web server which could be accessed on the World Wide Web at http://cast.engr.uic.edu/ aims to provide a detailed quantitative characterization of interior cavities and surface pockets of proteins, which are prominent concave regions of proteins that are frequently associated with binding events.

- Cavities are defined as buried unfilled empty space inside proteins after removing all heterogeneous atoms that are inaccessible to solvent molecules from outside.
- Pockets are defined as concave caverns with constrictions at the opening on the surface regions of proteins. Unlike cavities, pockets allow easy access for solvent molecules.

The CASTp server, in its most current release, was updated in 2006. The first version was launched in 1998 at the University of Minnesota. It is listed in the Research Consortium of Structural Bioinformatics PDB website at the San Diego Supercomputing Center (http://www.rcsb.org/pdb/) a central portal of structural bioinformatics worldwide. The server currently contains characterization of more than 2 million pockets and cavities that have been computed from PDB protein structures. The CASTp server is updated nightly with new PDB entries.

CASTp identifies all pockets and cavities on a protein structure and provides a detailed listing of all atoms participating in their formation. It also measures the volume and area of each pocket and cavity analytically, using both the solvent accessible surface and molecular surface models. In addition, it measures the size of mouth openings of individual pockets, which helps to assess the accessibility of binding sites to various ligands and substrates (Liang, 1998). The real advantages of the CAST algorithm are computation speed and robustness, rotational independence, precision of volume and area calculation, accurate pocket-cavity and mouth atom delineation.

3.5.1 Steps

- CASTp allows protein feature querying by insertion of the four letter PDB name of a protein structure or even structure can be uploaded.
- CASTp also offers on-line protein structure visualization through Chime or Mage.
- Once a protein is loaded in the browser, the user can interactively interrogate the protein structure.
- Summary information of measurement of individual pocket and cavity is displayed in a side list.
- Selection of a specific pocket from this list also presents the wall atoms forming the pocket in a separate small window.

These results can be obtained via email for further evaluation. Generally, results include: summary information of pockets and cavities, summary information of mouth openings of pockets, atomic synthesis of pocket walls, atomic synthesis of mouth rims and a RasMol and a KINEMAGE visualization script.

3.5.2 Computational method

CASTp utilizes computational geometry methods to analyze the geometrical features of proteins through their atomic coordinate data (Liang, 1998). CAST computation of pockets and cavities involves no human interactions. The only parameters required are the atomic van der Waals radii, and the radius of the probe sphere (usually 1.4 8, to approximate water). The probe sphere is used only to define molecular surface. All water molecules are first removed. The analytical area of each atom of the molecule, before and after removal of the ligand, is computed by an alpha-shape-based program VOLBL (Liang, 1998). Atoms showing a difference in area are designated as “in contact” with ligand. Pockets and voids are then computed for the ligand removed protein, and those containing any atoms “in contact” with the ligand are selected and defined as part of the ligand binding pocket. The initial stage of the computational procedure involves three steps: (a) the atomic radius for each atom of the PDB file is assigned using the utility program PDB2ALF; (b) the program DELCX is used to compute the three-dimensional weighted Delaunay triangulation, which takes into account the nonuniform nature of the atom radii; (c) the alpha shape or dual complex is computed using the program MKALF. The outputs of these three programs may then be utilized in three additional programs: For measurements of pockets and cavities, CAST accesses information computed by MKALF and computes the volume and area of pockets/cavities, as well as mouth area and circumference of pockets. The pocket information of all the selected proteins

were obtained using CASTp Server. While, the molecular visualizations of pockets and cavities were generated using PyMol.

3.6 Basic Local Alignment Search Tool (BLAST)

In bioinformatics, Basic Local Alignment Search Tool, or BLAST, is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences. A BLAST search enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold.

The BLAST program was designed by Eugene Myers, Stephen Altschul, Warren Gish, David J. Lipman and Webb Miller at the NIH and was published in J. Mol. Biol. in 1990. The BLAST program can either be downloaded and run as a command-line utility "blastall" or accessed for free over the web (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The BLAST program is based on an open-source format, giving everyone access to it and enabling them to have the ability to change the program code. Here query/input sequences are required in FASTA format or Genbank format while the BLAST output can be delivered in a variety of formats like HTML, plain text, and XML format.

BLAST works through use of a heuristic algorithm. Using this heuristic method, BLAST finds homologous sequences, not by comparing either sequence in its entirety, but rather by locating short matches between the two sequences. This process of finding initial words is called seeding. It is after this first match that BLAST begins to make local alignments. While attempting to find homology in sequences, sets of common letters, known as words are chosen. After making words for the sequence of interest, neighborhood words are also assembled. These words must satisfy a requirement of having a score of at least the threshold, T, when compared by using a scoring matrix (If Blastp search- BLOSUM62 scoring matrix is used). Once both words and neighborhood words are assembled and compiled, they are compared to the sequences in the database in order to find matches. The threshold score, T, determines whether a particular word will be included in the alignment or not. Each extension impacts the score of the alignment by either increasing or decreasing it. If the score is higher than a pre-determined Threshold value T, the alignment will be included in the results given by BLAST and in case score is lower than this pre-determined T, the alignment will cease to extend, preventing areas of poor alignment to be included in the BLAST results (Altschul, 1990).

There are now a handful of different BLAST programs available, which can be used depending on requirement and sequences. These different programs vary in query sequence input, the database being searched, and what is being compared. One of such program of our interest is BLASTp.

3.6.1 Protein-protein BLAST (BLASTp)

Standard protein-protein BLAST (BLASTp) is used for both identifying a query amino acid sequence and for finding similar sequences in protein databases. Like other BLAST programs, BLASTp is designed to find local regions of similarity. When sequence similarity spans the whole sequence, BLASTp also report a global alignment, which is the preferred result for protein identification purposes. Homologous sequences can be identified using the online NCBI protein-protein BLAST tool (Altschul, 1990). PDB sequence data were queried, using default parameters, against the non-redundant compilation of GenBank CDS translations, RefSeq proteins, PDB, SwissProt, PIR, and PRF protein sequences. While, homologous sequences were selected from each pathogenic Clostridium species and the single highest scoring sequence was selected from *Homo sapiens*. Sequence alignment used the Needleman and Wunsch algorithm with zero end gap penalties.

3.7 ConSurf

ConSurf is a web server, available at http://consurf.bioinfo.tau.ac.il/. The server is used for the identification of biologically important residues in protein sequences. Functionally important residues that take part in ligand binding and protein–protein interactions, are often evolutionarily conserved and are most likely to be solvent-accessible, whereas conserved residues within the protein core most probably have an important structural role in maintaining the protein's fold. Thus, estimated evolutionary rates, as well as relative solvent accessibility predictions, are assigned to each amino acid in the sequence; both are subsequently used to indicate residues that have potential structural or functional importance. In 2005, ConSurf-HSSP database was introduced which is a pre-calculated repository of ConSurf results based on multiple sequence alignments (MSAs) extracted from the HSSP database. Later on, ConSurf-DB was introduced, which present results in much more standard and cross platform formats. There are several ways to access the repository. For visual inspection of one or few proteins, a web interface, available at http://consurfdb.tau.ac.il/ supports 3D visualization and access to all supplementary data. The entire repository can be downloaded via ftp and used for large-scale automated studies.

For advanced uses, involving re-building of variants of the repository, the build scripts can be downloaded from the ConSurf-DB web site (Goldenberg, 2009).

ConSurf works by collecting the sequence homologues of each PDB entry using PSI-BLAST and align those using standard methods. The evolutionary conservation of each amino acid position in the alignment is calculated using the Rate4Site algorithm, implemented in the ConSurf web server. The algorithm takes into account the phylogenetic relations between the aligned proteins and the stochastic nature of the evolutionary process explicitly. Rate4Site assigns a conservation level for each position in the multiple sequence alignment using an empirical Bayesian inference. High evolutionary rate represents a variable position while low rate represents an evolutionarily conserved position. The conservation scores are normalized so that the average over all residues is zero, and the standard deviation is one. Low (negative) scores indicate the conserved positions while the high scores indicate the variable ones. The normalized scores are then binned into the 1–9 color codes representing the conservation grades, where 1 corresponds to maximal variability and 9 to maximal conservation. The buried or exposed predicted status of each residue is marked in the first row below the sequence. Slowly evolving (colour grades 8 and 9) and 'exposed' residues are predicted to be functional, whereas slowly evolving (colour grade 9) and 'buried' residues are predicted to be structurally important. Both are indicated in the second row below the sequence, as 'f' and's' respectively (Berezin, 2004). Visual inspection of the conservation patterns on the 3D structure enabled the identification of key residues that comprise the functionally important regions of the protein.

3.8 Pocket Scoring

Pockets were ranked by a scoring function with six terms, with lower scores indicating a better pocket (Nicola, 2008). The general equation of the scoring function is as given below. Here C_A is the absolute residue conservation and C_R is the relative residue conservation. The absolute and relative conservation measures were calculated from pair-wise alignments with homologous Clostridium spp. C_H is the absolute conservation with *Homo sapiens*. V and A are the pocket volume and surface area, respectively. Both V_0 and A_0 were set to 450, an ideal value for these variables. R is the resolution of the crystallographic structure.

$$S = -C_A - C_R + \frac{1}{1.0001 - C_H} + (V - V_0)^2 + (A - A_0)^2 + R$$

The Algorithm uses the following six terms: 1) absolute residue sequence identity 2) sequence identity relative to the protein surface 3) absolute sequence identity with the closest human homologue 4) pocket volume 5) pocket surface area and 6) crystallographic resolution of the protein. All of 71 pockets in different proteins of the *Clostridium perfringens* were ranked according to these criteria. For each pocket, sequence conservation statistics were gathered. Residue conservation was defined as (number of identical residues)/(total number of nearby residues). Relative conservation was defined as (residue conservation)/(solvent accessible residue conservation). Solvent accessible residues were those with at least 25% of their surface accessible to a water probe molecule. Finally, the relative contributions of the terms were balanced by scaling each so that the interquartile range (IQR) was 1 and the mean was 0.

This unique Pocket scoring approach incorporates accurate pocket identification along with sequence conservation with a similar organism, sequence conservation with the host and structure resolution. Sequence conservation between the *Clostridium perfringens* and similar organism was performed only on the pocket residues. Also the sequence conservation between each of the C. perfringens and their equivalent pockets in the most homologous human proteins was calculated to avoid the cross-reactivity with the host.

4. RESULT AND DISCUSSION

4.1 Protein Model Generation

Molecular models based on crystallographic PDB entries with released coordinates having high resolution, sequence identity closest to wild-type, and bound ligands were chosen with higher priority. And then corrections were applied using MolProbity to create the missing subunits as well as structures were optimized for the lowest energy orientation. Incorrectly formed HIS/GLN/ASN residues were also corrected. The list of selected PDB structure is given in **Table 4.1**.

4.2 Pocket Identification

For all the selected PDB ID's, pockets were identified using CASTp algorithm. All the pockets were then numbered according to their volume and pockets below a minimum volume were discarded. However, because several proteins contained repeated subunits in the quaternary structure, many pockets were replicated 2–6 times. As such, the number of pockets per protein varied greatly, from 1 to 24. For molecules with pockets containing prosthetic groups or cofactors, pocket identification was repeated with the inclusion of these bound molecules, and thus each complex was treated as a separate protein.

Of the 15 crystal structures of *Clostridium perfringens* obtained from the PDB, having a wide variety of functions, including ATP-binding, nucleotide biosynthesis, vesicular traffic, and catalytic activity, a total of 71 pockets were identified. These pockets are listed in **Table 4.2**, along with all the considered conservation parameters.

4.3 Pocket Scoring

The unique protocol of Pocket scoring incorporates six different variables in unique formula for identifying drug-like pockets. Here in order to select the most biologically significant targets and sites, this scoring formula has been used, which include terms for the pocket volume, size and shape (as determined by CASTp), as well as evolutionary conservation of residues lining each pocket (as determined by BLASTp and ConSurf web server). Finally the score was calculated for each pocket and the relative contributions of the terms were balanced by scaling each so that the interquartile range (IQR) was 1 and the mean was 0 (**Table 4.2**).

Protein names	**PDB ID**	**Resolution (Å)**
Phospholipase C (PLC)(Alpha-toxin)	1CA1	1.90
Perfringolysin O (Theta-toxin)	1PFO	2.20
Heat-labile enterotoxin B chain	2XH6	2.69
Hyaluronoglucosaminidase (Mu toxin)	2OZN	1.60
Iota toxin component Ia	1GIR	2.10
Epsilaon-toxin (Epsilon toxin)	1UYJ	2.60
Choloylglycine hydrolase (CBAH)	2BJG	2.10
Malonyl CoA-acyl carrier protein transacylase	3PTW	2.10
Phosphoenolpyruvate carboxylase (PEPCase)	3ODM	2.95
Energy-coupling factor transporter ATP-binding protein	3GFO	2.30
Phosphoribosylaminoimidazole-succinocarboxamide synthase (SAICAR synthetase)	3NUA	1.40
Geranyltranstransferase	3UCA	2.00
DHHA1 domain protein	3KEW	2.00
Putative uncharacterized protein	3O6U	2.50
Probable manganase-dependent inorganic pyrophosphatase	3L31	2.30

Table 4.1 *Clostridium perfringens* proteins whose structures were analyzed for the presence of pockets. The resolution of the structures ranged from 1.40 to 2.95 Å.

Table 4.2 List of all 71 Pockets of *Clostridium perfringens*

PDB ID	Pocket No.	Seq. Ident. (C_A)	Rel. Idnt. (C_R)	Hs. Ident. (H_S)	Vol. (V)	Area (A)	Res. (Å)	Score
1CA1	1	0.61	0.89	0.29	817.80	599.10	1.90	0.30589
1CA1	2	0.57	0.74	0.29	599.20	547.60	1.90	-1.00626
1CA1	3	0.29	0.37	0.29	290.70	288.80	1.90	0.682044
1CA1	4	0.30	0.50	0.10	201.10	213.50	1.90	5.304027
1CA1	5	0.54	0.60	0.18	192.30	145.90	1.90	39.72239
1CA1	6	0.55	0.71	0.33	139.10	126.70	1.90	1.952782
1CA1	7	0.22	0.66	0.22	118.40	169.10	1.90	-1.2118
1PFO	1	0.59	1.47	0.25	855.30	691.70	2.20	-0.22331
1PFO	2	0.69	1	0.12	762.20	517.10	2.20	4.249995
1PFO	3	0.47	0.57	0.12	637.30	295.80	2.20	5.039933
1PFO	4	0.55	0.90	0.20	444.80	393.40	2.20	-11.6695
1PFO	5	0.65	1.08	0.10	348.70	285.30	2.20	2.762141
1PFO	6	0.61	1.33	0.20	230.70	227.10	2.20	-11.2124
1PFO	7	0.54	0.66	0.18	128.10	151.70	2.20	40.56389
1PFO	8	0.60	1	0.10	131.50	114.60	2.20	5.480528
2XH6	1	0.27	0.28	0.11	413.20	300.70	2.69	6.05884
2XH6	2	0.33	0.49	0.11	308.20	318.50	2.69	5.723926
2OZN	1	0.92	1	0.71	228.70	136	1.60	0.120239
2OZN	2	0.92	1.09	0.71	151.70	125.80	1.60	0.597792
1GIR	1	0.50	1	0.11	401	411.50	2.10	2.973834
1GIR	2	0.20	0.29	0.25	403.90	429.60	2.10	-0.35797
1GIR	3	0.15	0.28	0.38	212.80	206.40	2.10	3.275694
1GIR	4	0.50	1.51	0.28	236	205.90	2.10	-0.82315
1GIR	5	0.25	0.50	0.16	138.40	114.60	2.10	12.29228
1GIR	6	0.70	1	0.10	118.70	189.80	2.10	4.472036
2BJG	1	0.36	0.42	0.29	257	177.80	2.10	1.591462
2BJG	2	0.30	0.34	0.29	224.80	179.20	2.10	2.059088
3PTW	1	0.87	2.02	0.70	486.50	366.10	2.10	-2.47852

3PTW	2	0.34	0.72	0.32	232.70	191.30	2.10	1.594975
3PTW	3	0.92	1	0.32	305.60	233.40	2.10	-1.19045
3PTW	4	0.10	0.13	0.30	164.50	170.30	2.10	3.606331
3PTW	5	0.54	0.60	0.44	140.80	117.40	2.10	3.114859
3ODM	1	0.29	0.38	0.24	391.90	423.70	2.95	0.15487
3ODM	2	0.14	0.19	0.24	514.30	300.20	2.95	1.224027
3ODM	3	0.12	0.16	0.15	577	311.80	2.95	10.08586
3ODM	4	0.28	0.48	0.15	276.40	201.30	2.95	9.921495
3ODM	5	0.14	0.14	0.14	268.10	127.10	2.95	10.3556
3ODM	6	0.25	0.30	0.14	257.90	180	2.95	9.370246
3ODM	7	0.34	0.40	0.14	313.70	156.20	2.95	8.904843
3ODM	8	0.17	0.25	0.14	205.10	212.30	2.95	9.745986
3ODM	9	0.36	0.65	0.10	223.30	166	2.95	6.865543
3ODM	10	0.25	0.40	0.14	365.70	375.70	2.95	7.834851
3GFO	1	0.88	2	0.50	444.70	315.10	2.30	-2.22469
3GFO	2	0.25	0.37	0.32	267	254.80	2.30	2.141879
3GFO	3	0.60	1.20	0.40	200.70	158.40	2.30	1.378634
3GFO	4	0.36	0.49	0.36	187.20	188.80	2.30	2.850066
3NUA	1	1	1.39	0.59	363.80	461.40	1.40	-3.15444
3NUA	2	0.63	2.52	0.13	268.20	231.50	1.40	1.007048
3NUA	3	0.45	0.84	0.27	452.80	208.10	1.40	-1.63203
3UCA	1	0.58	1.16	0.25	460.50	237.40	2.00	-2.38223
3UCA	2	0.60	3	0.20	183.30	173.10	2.00	-13.449
3UCA	3	0.60	1.50	0.20	148.60	142	2.00	-10.562
3UCA	4	0.90	1.36	0.40	120.50	146.40	2.00	0.456811
3UCA	5	0.60	0.90	0.40	117.30	142	2.00	2.130769
3KEW	1	0.50	0.79	0.23	862.40	558.30	2.00	-1.09537
3KEW	2	0.63	1.26	0.50	671.60	471.70	2.00	-0.44132
3KEW	3	0.88	1.25	0.47	437.20	298.70	2.00	-1.53483
3KEW	4	0.63	0.84	0.69	319.40	305.50	2.00	0.320638
3KEW	5	0.73	1.22	0.33	195.90	218.70	2.00	-0.38131
3KEW	6	0.44	0.57	0.22	255.90	125.80	2.00	-2.11279

3KEW	7	0.80	0.88	0.30	321	169.50	2.00	-0.64305
3KEW	8	0.33	0.66	0.25	125.80	136.20	2.00	1.296176
3KEW	9	0.57	0.80	0.14	126.60	104.80	2.00	7.622778
3KEW	10	0.33	0.42	0.11	136	137	2.00	7.512215
3O6U	1	0.12	1	0.18	342.90	273.40	2.50	39.70992
3O6U	2	0.34	0.40	0.25	162.80	100.10	2.50	2.5646
3O6U	3	0.20	0.66	0.20	116.60	136.70	2.50	-6.91512
3L31	1	0.69	2.23	0.44	536.70	323.20	2.30	-2.0993
3L31	2	0.87	1	0.13	266.40	265.50	2.30	4.002607
3L31	3	0.64	1.25	0.44	239.60	175.80	2.30	0.938776
3L31	4	0.22	1	0.11	109.10	143.90	2.30	6.97275

Table 4.2 List of all 71 pockets of 15 different proteins of *Clostridium perfringens*

There exists an unmet need to successfully identify novel targets in important pathogens. So, here we report a study that focuses on the identification of small molecule binding sites, which could serve as novel targets. For, this purpose, we have used CASTp pocket identifying algorithm and a unique pocket scoring algorithm, which incorporates six different variables for calculation and scoring. Firstly, for pocket identification, there are different softwares which can be used for pocket identification; some uses geometric methods while other uses energy computation method. So, method used by us, i.e. CASTp is based on geometric computation and it has showed ~71-75% success rate in accurately identifying the pockets. Secondly, the scoring formula used, includes six different variables for calculation, two of which, pocket volume and area were determined by CASTp and third term protein resolution was determined by PDB information and rest of the three terms (Absolute residue conservation, relative residue conservation and conservation with human sequence) were determined using BLASTp and ConSurf web server information. These residue conservation parameters provide essential evolutionary conservation properties of the proteins residues lining the pockets.

A high conservation between proteins in the *Clostridium perfringens* and homologous proteins from a related species can be considered as an advantage in the process of pocket scoring. Because, this ensures the functional relevance and also increases the chances that drugs developed to target these pockets will cross-react with different members of the Clostridium genus. However, it is important to mention that this is a very limited attempt to select for the most functionally relevant proteins in a genome. For optimal evaluation, this approach needs to be coupled with other biological and functional aspects of each protein. Another sequence analysis used in the final pocket scoring function includes the comparison between *C. perfringens* and *H. sapiens*. Conservation of the residues between each putative target and homologous proteins in human could potentially lead to toxicity due to cross-reactivity, and thus considered unfavorable. While this measure does not account for all types of potential cross-reactivity in human, it enriches the set of top-scoring targets with those that are more biologically significant.

All the pockets identified were finally ranked according to scoring formula, with more negative score showing best binding site. So, here we report few pockets with good score, which may serve as novel targets in **Table 4.3**.

PDB ID	Pocket No.	Seq. Ident. (C_A)	Rel. Ident. (C_R)	Hs. Ident. (H_S)	Vol. (V)	Area (A)	Res. (Å)	Score
3NUA	1	1	1.39	0.59	363.80	461.40	1.40	-3.15
3PTW	1	0.87	2.02	0.70	486.50	366.10	2.10	-2.88
3UCA	1	0.58	1.16	0.25	460.50	237.40	2.00	-2.38
3GFO	**1**	**0.88**	**2**	**0.50**	**444.70**	**315.10**	**2.30**	**-2.22**
3KEW	6	0.44	0.57	0.22	255.90	125.80	2.00	-2.11
3L31	1	0.69	2.23	0.44	536.70	323.20	2.30	-2.09

Table 4.3 Pockets with best score, which may serve as good target

These six proteins have been found to be involved in various metabolic activities of Clostridium, like in purine biosynthesis, nucleotide binding, ATP-binding and isoprenoid biosynthesis process. Also, four of these proteins, namely 3UCA (Geranyltranstransferase), 3L31 (Probable manganase-dependent inorganic pyrophosphatase), 3GFO (Energy-coupling factor transporter ATP-binding protein) and 3NUA (Phosphoribosylaminoimidazole-succinocarboxamide) have been reported as essential protein for *Clostridium perfringens* (Chhabra, 2010). From the ConSurf web-server information, we observed that some of the amino acid residues lining the pocket cavity were highly conserved among the similar organism (**Fig 4.1-4.4**). Highly conserved region, in theory, are more likely to be physiologically important; however, they need to be experimentally validated. Further, a pocket can be stated as ideal, if it is of right size, so as to accommodate drug-sized molecule and slightly buried, which increases interaction surface area and finally the amino acid forming the pocket should not be too polar, so as to allow drug-like properties in ligands (as shown in **Fig 4.5-4.10**).

Fig 4.1 ConSurf web-server results for 3NUA PDB structure showing evolutionary conserved regions as well as residues which are exposed or buried.

(The conservation scale:

1 2 3 4 5 6 7 8 9

Variable **Average** **Conserved**

e - An exposed residue according to the neural-network algorithm.

b - A buried residue according to the neural-network algorithm.

f - A predicted functional residue (highly conserved and exposed).

s - A predicted structural residue (highly conserved and buried).

Residues underlined red, are pocket lining residues found to be highly conserved, functional or structural amino acid.)

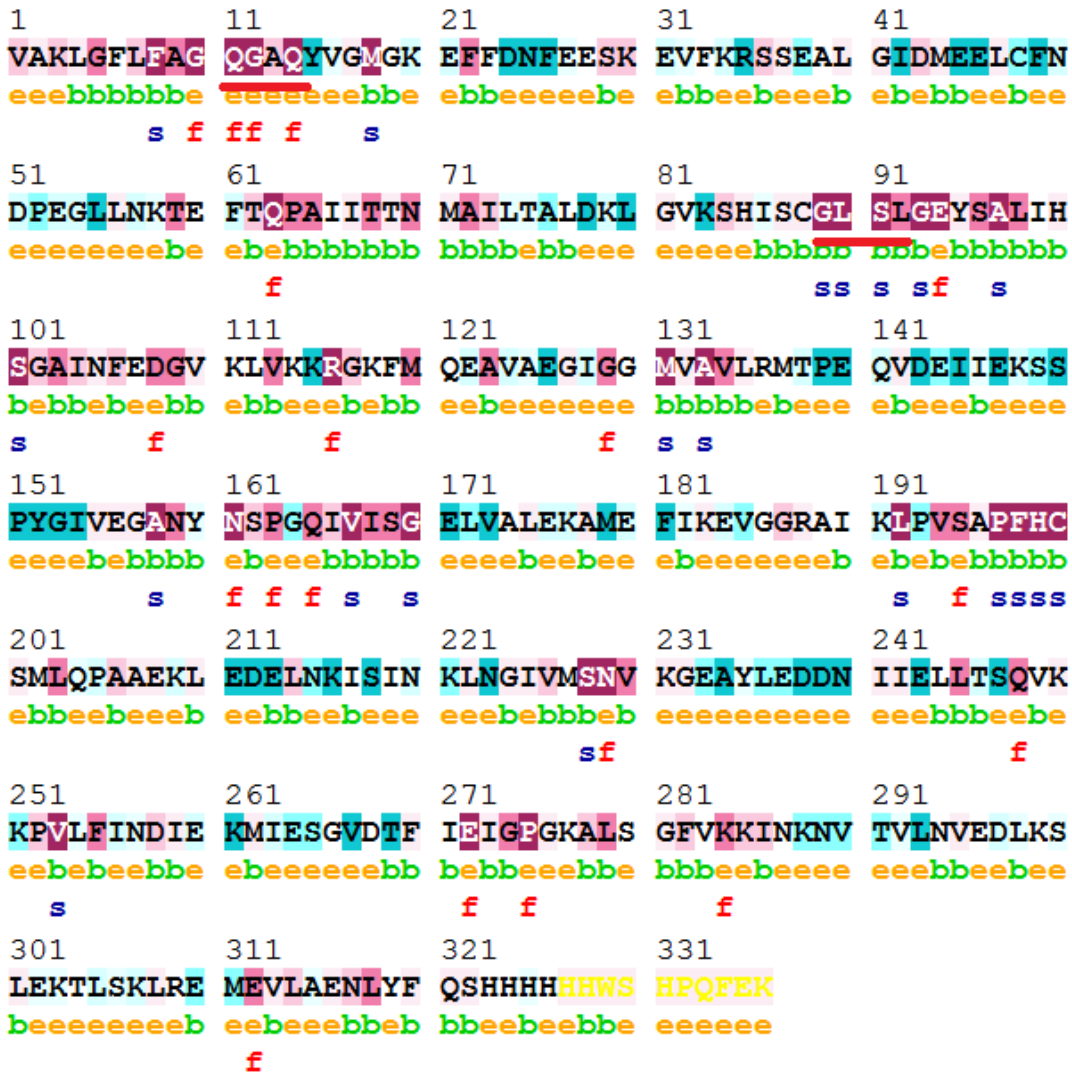


Fig 4.2 ConSurf web-server results for 3PTW PDB structure showing evolutionary conserved regions as well as residues which are exposed or buried.

(The conservation scale:

e - An exposed residue according to the neural-network algorithm.

b - A buried residue according to the neural-network algorithm.

f - A predicted functional residue (highly conserved and exposed).

s - A predicted structural residue (highly conserved and buried).

Residues underlined red, are pocket lining residues found to be highly conserved.)

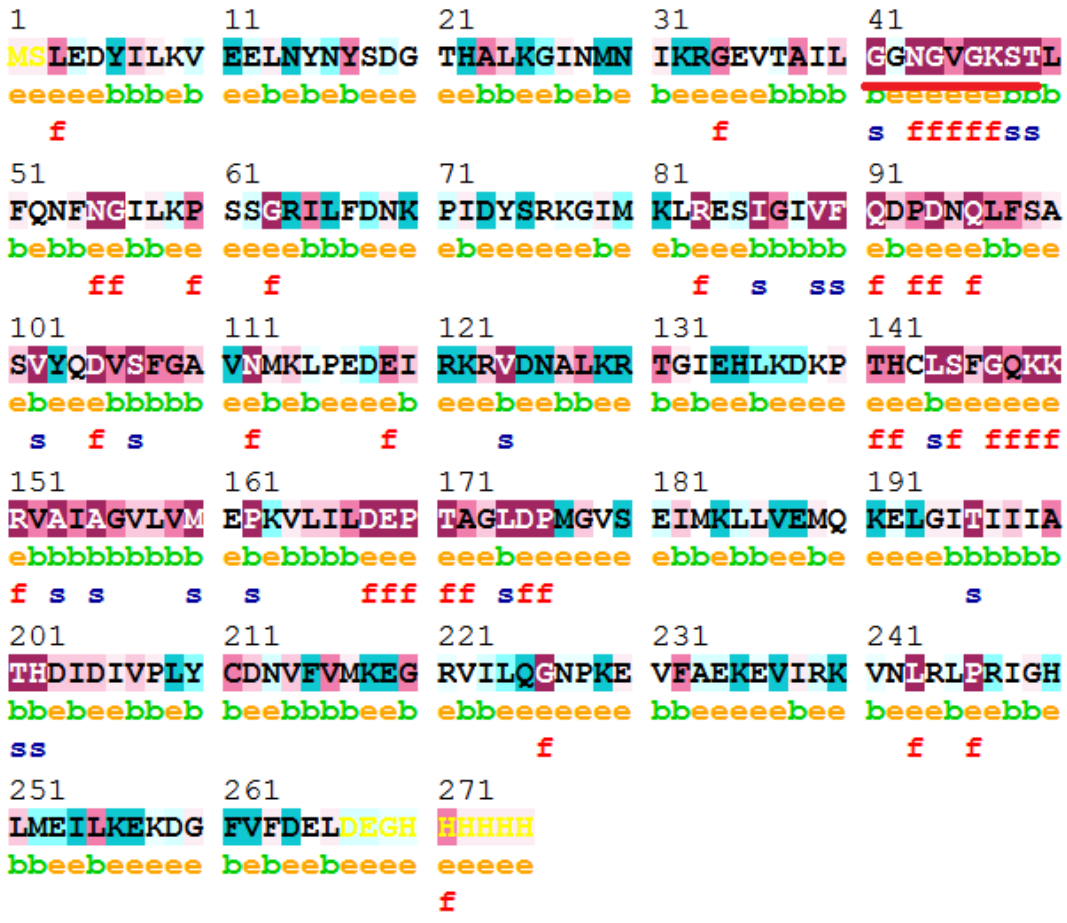


Fig 4.3 ConSurf web-server results for 3GFO PDB structure showing evolutionary conserved regions as well as residues which are exposed or buried.

(The conservation scale:

1 2 3 4 5 6 7 8 9

Variable Average Conserved

e - An exposed residue according to the neural-network algorithm.

b - A buried residue according to the neural-network algorithm.

f - A predicted functional residue (highly conserved and exposed).

s - A predicted structural residue (highly conserved and buried).

Residues underlined red, are pocket lining residues found to be highly conserved, functional or structural amino acid.)

Fig 4.4 ConSurf web-server results for 3UCA PDB structure showing evolutionary conserved regions as well as residues which are exposed or buried.

(The conservation scale:

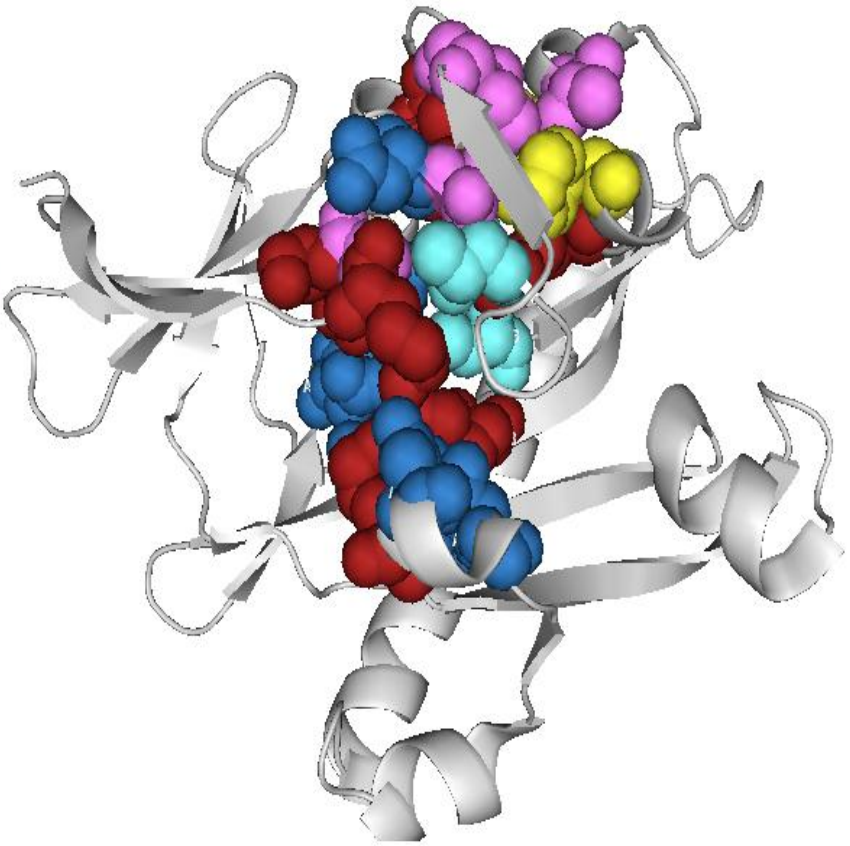
e - An exposed residue according to the neural-network algorithm.

b - A buried residue according to the neural-network algorithm.

f - A predicted functional residue (highly conserved and exposed).

s - A predicted structural residue (highly conserved and buried).

Fig 4.5 PDB structure 3NUA
(Pocket forming residues shown as spheres-
Positive-RED (K,H,R)
Negative-BLUE (D,E)
Non-Polar- CYAN (S,T,C,N,Q)
Hydrophobic-VIOLET (G,A,V,L,I,M,P)
Aromatic-YELLOW (F,Y,W)



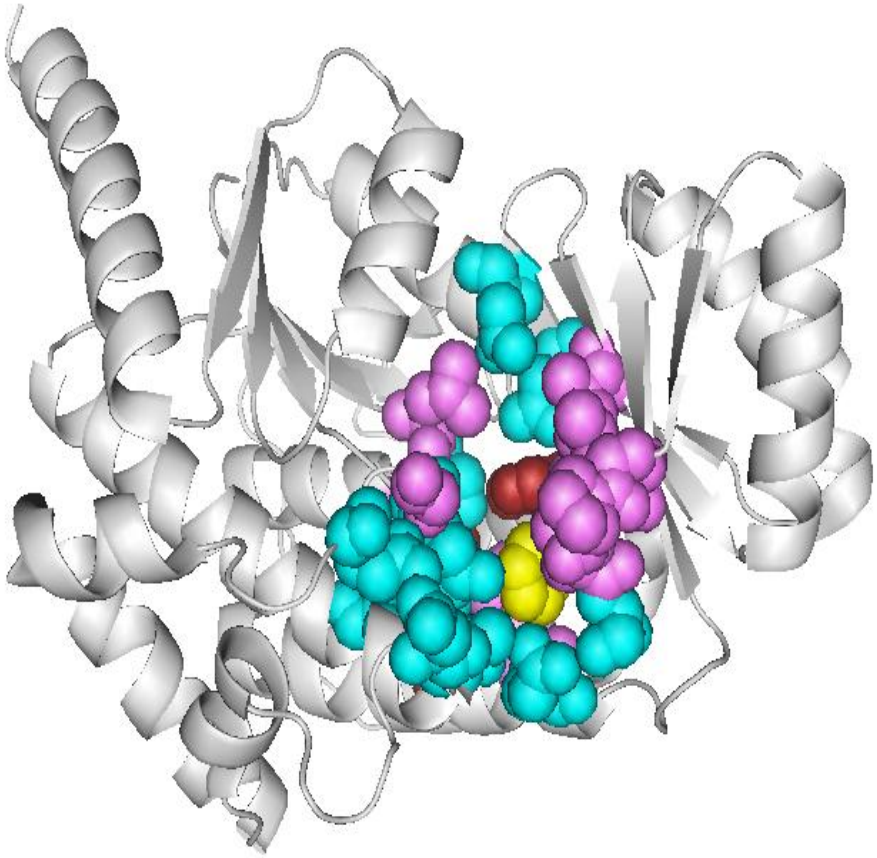


Fig 4.6 PDB structue 3PTW
(Pocket forming residues shown as spheres-

Positive- RED (K,H,R)
Negative- BLUE (D,E)
Non-Polar- CYAN (S,T,C,N,Q)
Hydrophobic- VIOLET (G,A,V,L,I,M,P)
Aromatic- YELLOW (F,Y,W))

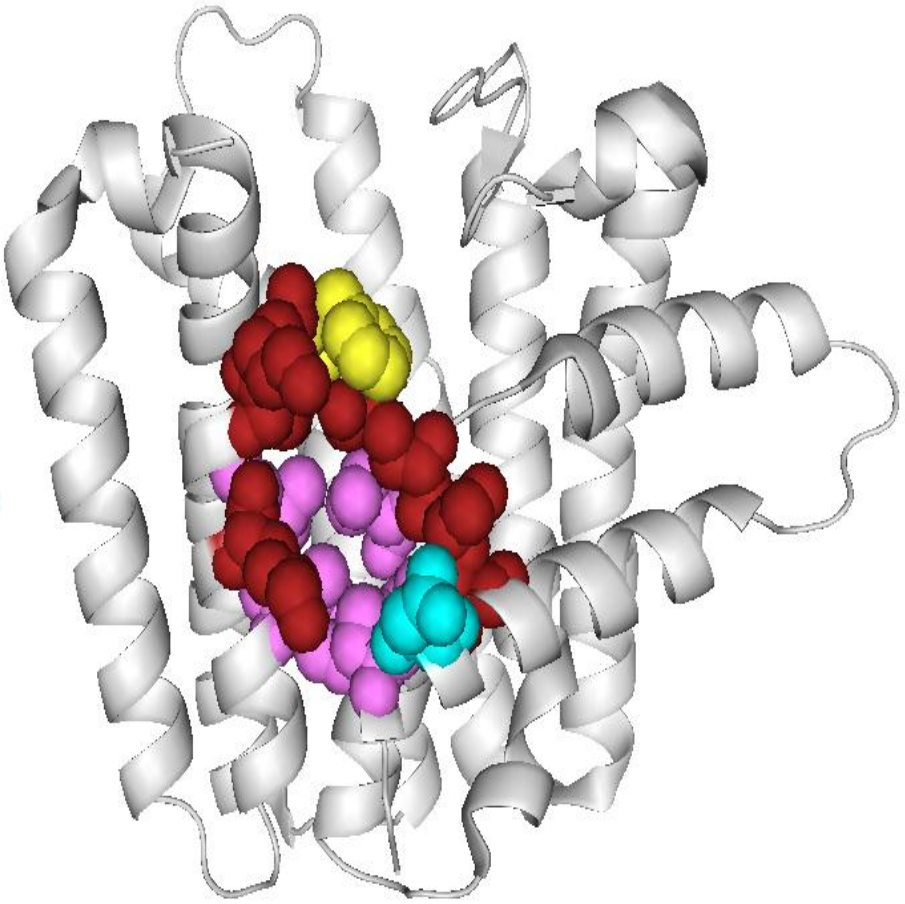


Fig 4.7 PDB structure 3UCA

(Pocket forming residues shown as spheres-

Positive- RED (K,H,R)
Negative- BLUE (D,E)
Non-Polar- CYAN (S,T,C,N,Q)
Hydrophobic- VIOLET (G,A,V,L,I,M,P)
Aromatic- YELLOW (F,Y,W))

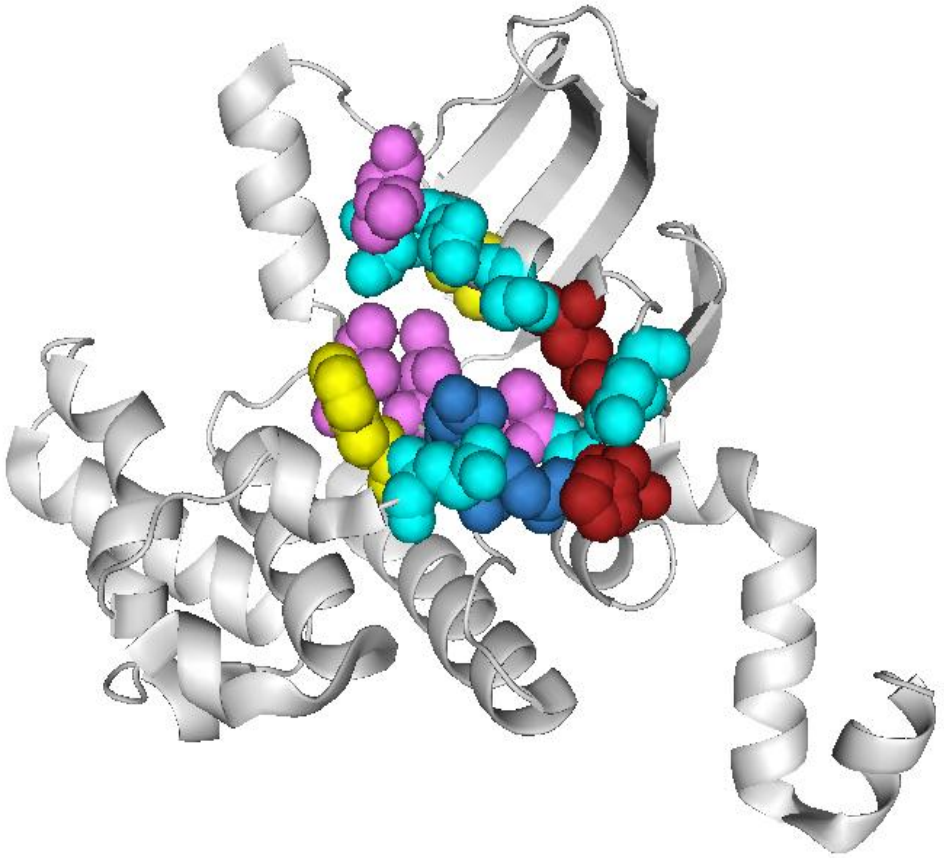


Fig 4.8 PDB structure 3GFO

(Pocket forming residues shown as spheres-

Positive- RED (K,H,R)
Negative- BLUE (D,E)
Non-Polar- CYAN (S,T,C,N,Q)
Hydrophobic- VIOLET (G,A,V,L,I,M,P)
Aromatic- YELLOW (F,Y,W))

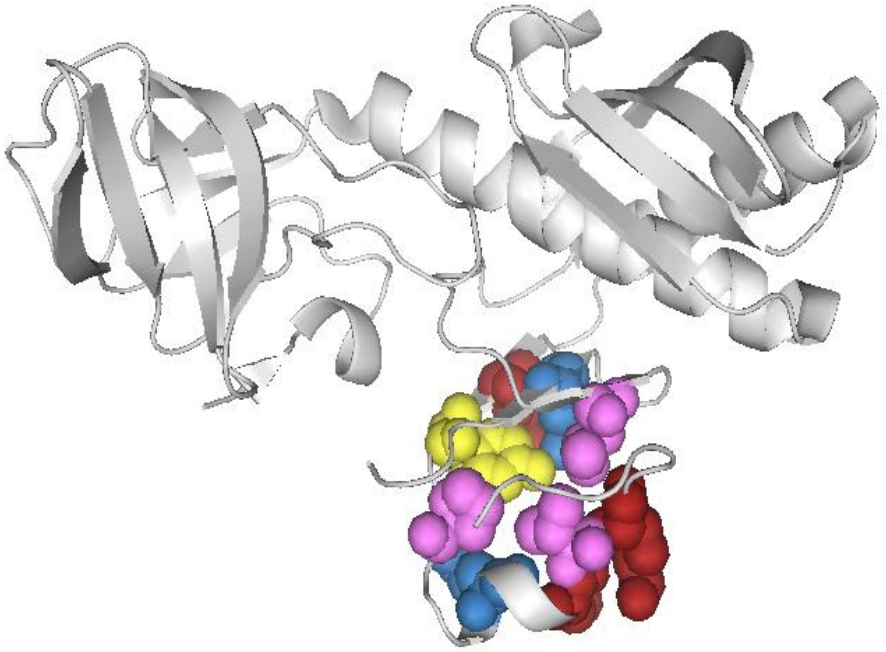


Fig 4.9 PDB structure 3KEW

(Pocket forming residues shown as spheres-

Positive- RED (K,H,R)
Negative- BLUE (D,E)
Non-Polar- CYAN (S,T,C,N,Q)
Hydrophobic- VIOLET (G,A,V,L,I,M,P)
Aromatic- YELLOW (F,Y,W))

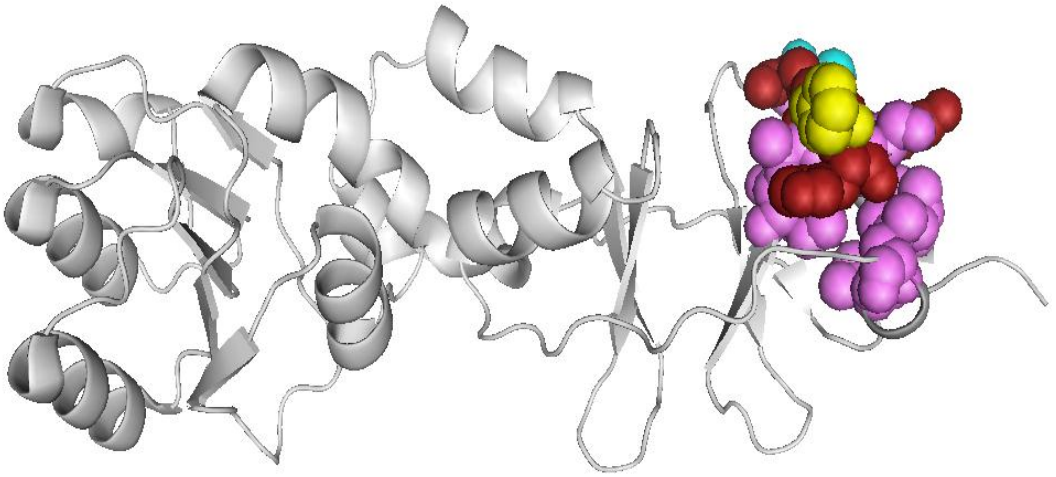


Fig 4.10 PDB structure 3L31

(Pocket forming residues shown as spheres-

Positive- RED (K,H,R)
Negative- BLUE (D,E)
Non-Polar- CYAN (S,T,C,N,Q)
Hydrophobic- VIOLET (G,A,V,L,I,M,P)
Aromatic- YELLOW (F,Y,W))

Amongst these six pockets, studies on 3GFO have been done by other scientists too. 3GFO, an ABC transporter is member of ubiquitously present ATP dependent transmembrane solute pumps and ion channels. These transporter superfamilies contain both uptake and efflux transport systems and form one of the largest transporters. 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. 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.

Similarly, so as to validate our results we further need to design small molecules which can serve as ligand molecule and to study how they interact with the protein, particularly the pocket forming residues. As a matter of fact, we already know that a PDB structure 3GFO has a SO_4 and 3NUA has ADP as ligand, so in **Fig 4.11-4.12** shows, how they interact with protein structure. In case of ATP transporter protein, (3GFO), we observed that pocket forming residues ASN41, LYS45, SER46 are clearly involved in interacting with ligand molecule. And in case of 3NUA, we observed pocket forming residues, LYS12, ALA13, LYS14, LYS124, GLU180, ASP192 are involved in interactions.

The pocket forming amino acid interaction and conservation studies suggest that further success can be achieved, by designing small ligand molecule which could serve as inhibitors. So, this pocket scoring hypothesis needs to be confirmed by the valid drug design approach. However, already explored ATP transporter protein information could serve as an excellent starting point for further studies, as well as all other best scoring pockets can be utilized for the development of potential inhibitors/modulators.

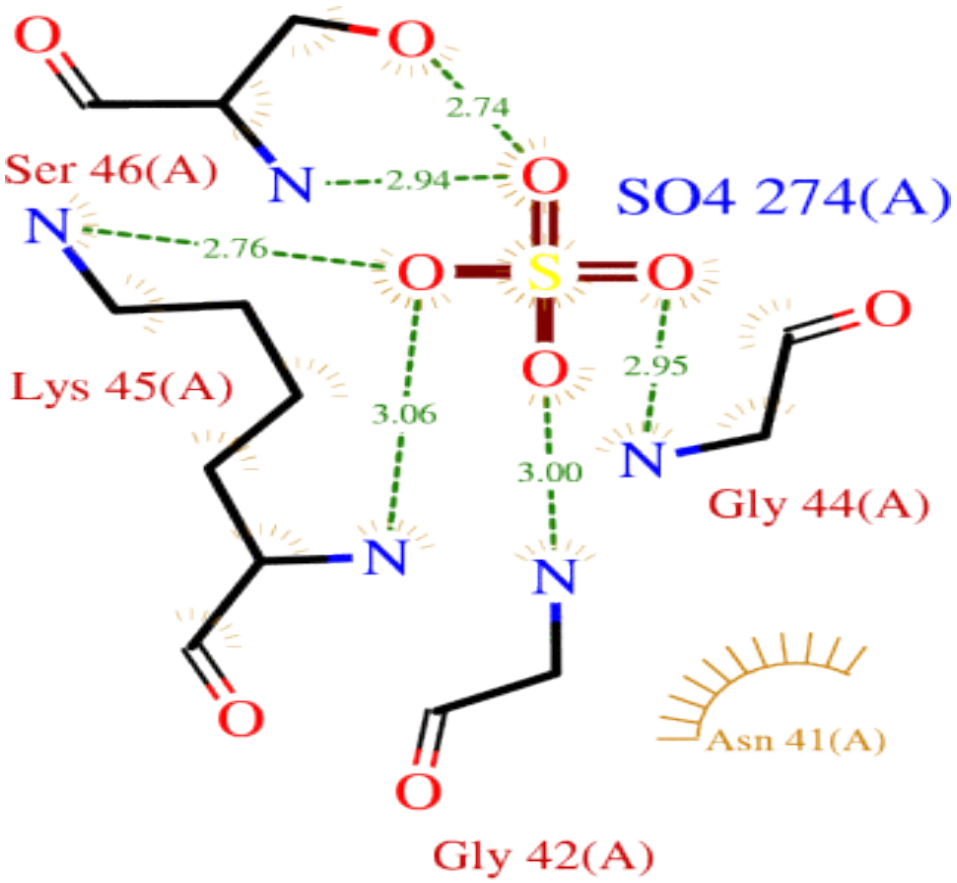


Fig 4.11 Automatically generated schematic diagram of protein-ligand interaction for PDB structure 3GFO (Wallace, 1996)**.**

(The interactions shown are those mediated by hydrogen bonds and by hydrophobic contacts. Hydrogen bonds are indicated by dashed lines between the atoms involved, while hydrophobic contacts are represented by an arc with spokes radiating towards the ligand atoms they contact. The contacted atoms are shown with spokes radiating back.)

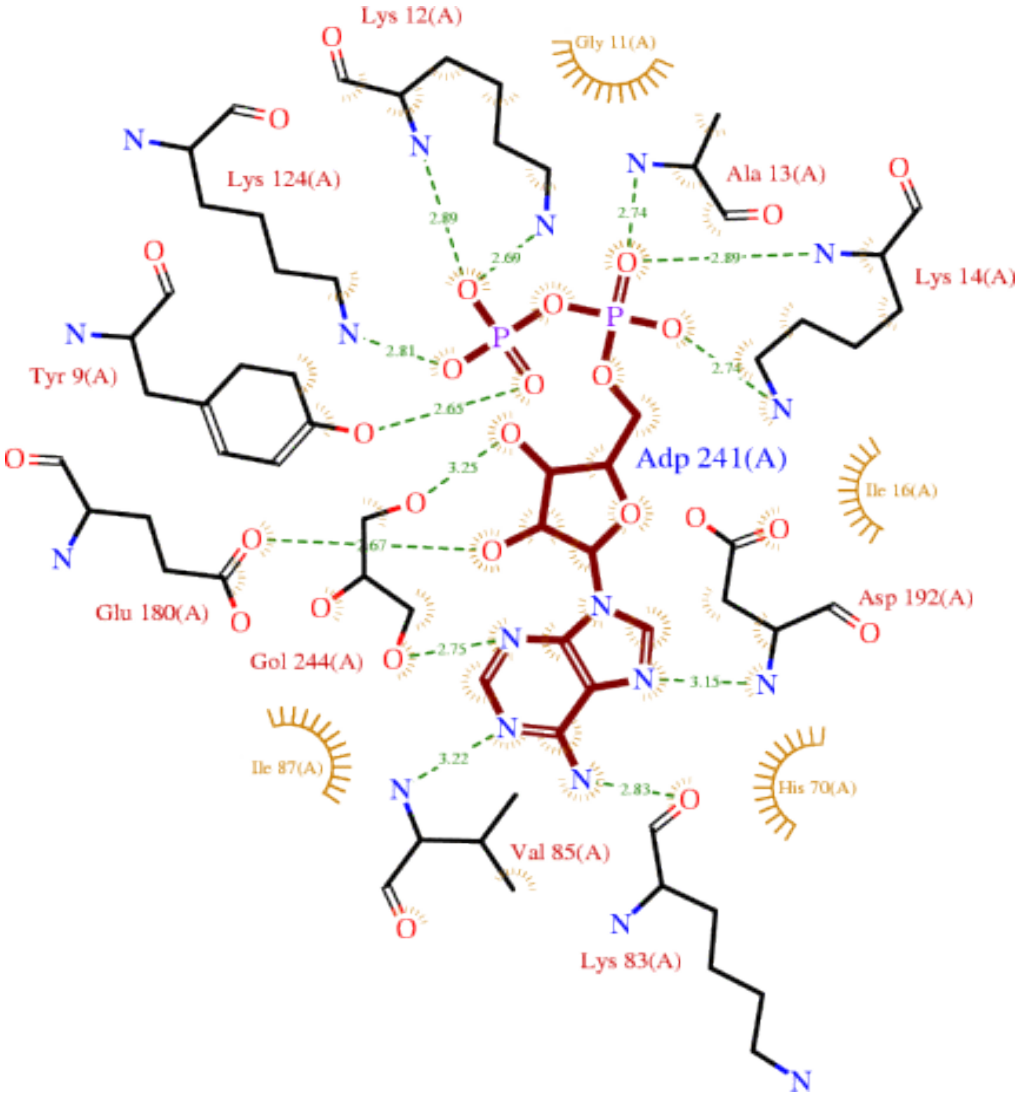


Fig 4.12 Automatically generated schematic diagram of protein-ligand interaction for PDB structure 3NUA (Wallace, 1996)**.**

(The interactions shown are those mediated by hydrogen bonds and by hydrophobic contacts. Hydrogen bonds are indicated by dashed lines between the atoms involved, while hydrophobic contacts are represented by an arc with spokes radiating towards the ligand atoms they contact. The contacted atoms are shown with spokes radiating back.)

5. CONCLUSION AND FUTURE PERSPECTIVES

Clostridium perfringens, is a very important pathogen, which causes many gastrointestinal illnesses with severity ranging from necrotizing enterocolitis in infants, food poisoning to gas gangrene (clostridial myonecrosis) in adults. So, here we report a study that focuses on methodology for assessing the Drugability of all small-molecule binding pockets in a pathogen *Clostridium perfringens*. Usually the essential genes and their protein products serve as good target. Essential genes are by definition necessary for replication and viability, and therefore the deletion, interruption or blocking of the protein expressed by an essential gene, results in death of the organism, making them attractive targets for drug development. By identifying a protein essential to a bacteria and developing inhibitors to that protein, new antimicrobials can be identified. Essential genes that are conserved across bacterial genera have been proposed as promising candidates for broad spectrum drug targets, active against multiple bacterial species.

For this study molecular models based on crystallographic PDB entries with released coordinates with high resolution were utilized for accurate pocket identification using geometric computation based CASTp server. The identified pockets were scored and ranked using the unique protocol of Pocket scoring which incorporates six different variables in formula for identifying drug-like pockets. Here in order to select the most biologically significant targets and sites, this scoring formula has been used, which include terms for the pocket volume, size and shape (as determined by CASTp), as well as evolutionary conservation of residues lining each pocket (as determined by BLASTp and ConSurf web server). A high conservation between proteins in the *Clostridium perfringens* and homologous proteins from a related species can be considered as an advantage in the process of pocket scoring. Because, this ensures the functional relevance and also increases the chances that drugs developed to target these pockets will cross-react with different members of the Clostridium genus. However, it is important to mention that this is a very limited attempt to select for the most functionally relevant proteins in a genome. For optimal evaluation, this approach needs to be coupled with other biological and functional aspects of each protein.

Finally using this study, few potential pockets were identified. These proteins were found to be involved in various metabolic activities of Clostridium, like in purine biosynthesis, nucleotide binding, ATP-binding and isoprenoid biosynthesis process. From the ConSurf web-server information, we observed that some of the amino acid residues lining the pocket

cavity were highly conserved among the similar organism. Highly conserved region, in theory, are more likely to be physiologically important; however, they need to be experimentally validated. Further, a pocket can be stated as ideal, if it is of right size, so as to accommodate drug-sized molecule and slightly buried, which increases interaction surface area and finally the amino acid forming the pocket should not be too polar, so as to allow drug-like properties in ligands.

Amongst the six potential pockets, studies done on PDB ID 3GFO were found to be more interesting. 3GFO, an ABC transporter is member of ubiquitously present ATP dependent transmembrane solute pumps and ion channels. The ABC transporters couple hydrolysis of ATP to the translocation of various substrates across cell membranes. 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. 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. However, to validate our results we further need to design small molecules which can serve as ligand molecule and to study how they interact with the protein, particularly the pocket forming residues. So, this pocket scoring hypothesis needs to be confirmed by the valid drug design approach. Although, already explored ATP transporter protein information could serve as an excellent starting point for further studies, as well as all other best scoring pockets can be utilized for the development of potential inhibitors/modulators.

6. REFERENCES

1. Altschul, S. G. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215:403–410.

2. Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S. (2004). UniProt: The Universal Protein knowledgebase. Nucleic Acids Research , 32: 115D–1119.

3. Berezin, C. G. (2004). ConSeq: the identification of functionally and structurally important residues in protein sequnces. Bioinformatics, 1322-1324.

4. Bernstein, F. K. (1977). The protein data bank: A computer based archival file for macromolecular structures. Journal of Molecular Biology, 112:535.

5. Chhabra, G. S. (2010). Identification and modelling of a drug target for *Clostridium perfringens* SM101. Bioinformation, 278-289.

6. Cooper, I. A. (2011). The in silico prediction of bacterial essential genes. Science against microbial pathogens: communicating current research and technological advances.

7. Czeczulin, J. C. (1996). Regulated Expression of *Clostridium perfringens* Enterotoxin in Naturally cpe-Negative Type A, B, and C Isolates of *C. perfringens*. Infection and Immunity, 64(8): 3301–3309.

8. Damme, V. J. (1996). Analysis of the catalytic site of the actin ADP-ribosylating *Clostridium perfringens* iota toxin. Federation of European Biochemical Societies Letters, 19;380(3):291-5.

9. Dundas, J. O. (2006). CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acid Research, W116-W118.

10. Franco, B. M. (2009). The determinants of the antibiotic resistance process. Infection and Drug Resistance, 2, 1-11.

11. Gao, Z. A. (2011). Use of *Clostridium perfringens* Enterotoxin and the Enterotoxin Receptor-Binding Domain (C-CPE) for Cancer Treatment: Opportunities and Challenges. Journal of Toxicology, Volume 2012, Article ID 981626,.

12. Goldenberg, O. E. (2009). The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. Nucleic Acid Research, D323-D327.

13. Grosdidier, S. T. R. (2009). Computer applications for prediction of protein–protein interactions and rational drug design. Advances and Applications in Bioinformatics and Chemistry, 2, 101–123.

14. Liang, J. E. (1998). Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. Protein Science, 7;1884-1897.

15. Matsushita, O. Y. (1994). Purification and Characterization of a *Clostridium perfringens* 120-Kilodalton Collagenase and Nucleotide Sequence of the Corresponding Gene. Journal of Bacteriology, 176(1): 149-156.

16. Miyamoto, K. L. (2008). Sequencing and Diversity Analyses Reveal Extensive Similarities between Some Epsilon-Toxin-Encoding Plasmids and the pCPF5603 *Clostridium perfringens* Enterotoxin Plasmid. Journal of Bacteriology, Vol. 190, No. 21, 7178–7188.

17. Moellering, R. (2011). Discovering new antimicrobial agents. Iternational Journal of Antimicrobial Agents, 37:2-9.

18. Nicola, G. S. (2008). New Method for the Assessment of All Drug-Like Pockets Across a structural genome. Journal of Computational Biology, 15(3): 231-240.

19. Perelle, S. G. (1993). Characterization of *Clostridium perfringens* Iota-Toxin Genes and Expression in *Eschenichia coli*. Infection and Immunity, Vol. 61, No. 12, p. 5147-5156.

20. Piddock, L. R. (2000). Activity of antibiotics used in human medicine for *Campylobacter jejuni* isolated from farm animals. Journal of Antimicrobial Chemotherapy, 46:303-6.

21. Randy, J. P. (2011). A new generation of crystallographic validation tools for the protein data bank. Structure, 1395-1412.

22. Rood, J. A. (1991). Molecular Genetics and Pathogenesis of *Clostridium perfringens*. Microbiological Reviews, 621-648.

23. Ryan, K. R. (2004). Sherris Medical Microbiology, 4th edition. McGraw Hill.

24. Sakurai, J. N. (2004). Clostridium perfringens alpha toxin: characterization and mode of action. Journal of Biochemistry, 569-574.

25. Shepard, L. S. (2000). The mechanism of pore assembly for a cholesterol-dependent cytolysin: formation of a large prepore complex precedes the insertion of the transmembrane beta-hairpins. Biochemistry, 10284-93.

26. Shimizu, T. O. (2002). Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. PNAS, vol. 99, no. 2.

27. Tansuphasiri, U. M. (2005). Antimicrobial resistance among *Clostridium perfringens* isolated from various sources. University of Thialand, Vol. 36 No.4.

28. Vincent, B. A. (2010). MolProbity: all atom structure validation for macromolecular crystallography. Acta Crystallography, D66,12-21.

29. Wallace, A. L. (1996). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Engineering, 127-134.

30. Weisel, M. P. (2007). PocketPicker: analysis of ligand binding-sites with shape descriptors. Chemistry Central Journal, 1:7.

31. Wells, C. W. (1996). Clostridia: Sporeforming Anaerobic Bacilli in: Baron's Medical Microbiology, 4th edition. University of Texas Medical Branch.

7. Appendix

Websites and Softwares URL:

1. UniProtKB- http://www.uniprot.org
2. BLASTp- http://blast.ncbi.nlm.nih.gov/Blast.cgi
3. PDB- http://www.rcsb.org/pdb
4. PyMOL- http://www.PyMol.org
5. MolProbity- http://molprobity.biochem.duke.edu.
6. CASTp- http://cast.engr.uic.edu/
7. ConSurf - http://consurf.bioinfo.tau.ac.il/