Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

BIOLOGIA CELLULARE, MOLECOLARE E INDUSTRIALE

Progetto 2: Biologia Funzionale e Molecolare

Ciclo XXV

Settore Concorsuale di afferenza: 05/E2

Settore scientifico-disciplinare: BIO/11

Diversity of *cwp* loci in clinical isolates of *Clostridium difficile*

Presentata da: **Manuele Biazzo**

Coordinatore Dottorato

Chiar.mo Prof. **Vincenzo Scarlato**

Relatori

Dr. **Cesira L. Galeotti**

Chiar.mo Prof. **Vincenzo Scarlato**

Esame finale anno 2013

TABLE OF CONTENTS

ATTIVITÀ DI RICERCA

Durante il Dottorato di Ricerca mi sono occupato dello studio di "cell wall proteins" (Cwp) di *Clostridium difficile* utilizzando due diversi approcci. In particolare ho determinato la presenza, la sequenza e la variabilità d'espressione di antigeni identificati come putative Cwp tramite un approccio di Reverse Vaccinology. Inoltre, mediante la tecnologia del protein microarray, ho analizzato se tali proteine sono esposte sulla superficie batterica durante l'infezione e capaci di indurre una risposta immunitaria in hamsters infetti da *C. difficile*. Tra le tecnologie utilizzate nel mio studio rientrano : clonaggio, espressione e purificazione di proteine antigeniche di *C. difficile*, l'utilizzo di protein microarrays per l'analisi di sieri umani e di animali, la manipolazione di isolati clinici di C. difficile e studi di epidemiologia molecolare tramite l'utilizzo di software specifici.

Nel periodo del Dottorato di Ricerca sono stato co-autore del seguente lavoro scientifico e di due elaborati relativi a congressi :

Biazzo M. *et al.* "Diversity of *cwp* loci in clinical isolates of *Clostridium difficile*" (Submitted to Journal of Medical Microbiology, 2013).

Biazzo M. *et al.* Poster: "Diversity of *cwp* loci in clinical isolates of *Clostridium difficile*". 3rd International *Clostridium difficile* Symposium, Bled – Slovenia 22 to 24 September 2010.

Biazzo M. *et al.* Poster: "Diversity of *cwp* loci in clinical isolates of *Clostridium difficile*". 4th International *Clostridium difficile* Symposium, Bled – Slovenia 20 to 22 September 2012.

ABSTRACT

An increased incidence of *Clostridium difficile* infection (CDI) is associated with the emergence of epidemic strains characterised by high genetic diversity. Among the factors that may have a role in CDI there is a family of 29 paralogs, the cell wall proteins (CWPs), which compose the outer layer of the bacterial cell and are likely to be involved in colonisation. Previous studies have shown that 12 of the29 *cwp* genes are clustered in the same region, named after *slpA* (*cwp1*) the *slpA* locus, whereas the remaining 17 paralogs are distributed throughout the genome.

The variability of 14 of these 17 *cwp* paralogs was determined in 40 *C. difficile* clinical isolates belonging to six of the currently prevailing PCR ribotypes. Based on sequence conservation, these *cwp* genes were divided into two groups, one comprising *cwp* loci having highly conserved sequences in all isolates, and the other 5 loci showing low genetic conservation between isolates of the same PCR ribotype as well as between different PCR ribotypes. Three conserved CWPs, Cwp16, Cwp18 and Cwp25, and two variable ones, Cwp26 and Cwp27, were characterised further by Western blot analysis of total cell extracts or S-layer preparations of the *C. difficile* clinical isolates. Expression of genetically invariable CWPs is well conserved in all isolates, while genetically variable CWPs are not always expressed at comparable levels even in strains containing identical sequences but belonging to different PCR ribotypes.

In addition, we chose to analyse the immune response obtained in a protection experiment, carried out in hamsters, using a protein microarray approach to study the *in vivo* expression and the immunoreactivity of several surface proteins, including 18 Cwps.

1. INTRODUCTION

1.1 Clostridium difficile : general features

Clostridium difficile is a spore-forming, Gram-positive, obligate anaerobic bacterium and is the most common cause of nosocomial infectious diarrhea. It is found in the commensal flora of 3% of the adults and 24 % of patients in hospitals (Gould and McDonald 2008). Since the 1970s this species is recognized as a cause of human gastrointestinal infection, and now is known to cause the most frequent healthcare-acquired infectious diarrhea in developed countries (Voth and Ballard 2005; Barbut, Gariazzo et al. 2007).

C. difficile infection (CDI) shows different clinical symptoms, ranging from uncomplicated asymptomatic carriage and mild diarrhea to life-threatening toxic megacolon and pseudomembranous colitis (PMC) requiring surgical intervention (Barbut, Gariazzo et al. 2007). Over the last decade the incidence of *C. difficile* infection has dramatically increased due to the emergence of new lineages (such as PCR ribotypes 027, 017 and, more recently, 078) that are more transmissible and cause more severe infections. A thorough analysis of strains from different sources and geographical regions shows significant microdiversity of clonal complexes demonstrating the evolution of *C. difficile* (Cairns, Stabler et al. 2012).

The study of this pathogen takes into account a range of selective pressures created by human activity and practices in healthcare settings. Typical features of CDI include: watery diarrhea, abdominal pain and cramps, lower quadrant tenderness, fever, leukocytosis and hypoalbuminemia. Re-occurrences can arise in 20% of the cases after the first episode and in 50% after the second episode even after treatment (Barbut, Gariazzo et al. 2007).

Infection begins with the ingestion of *C. difficile* spores excreted from feces of infected patients. Spores are easily transmitted via persons (usually hands of healthcare staff), fomites and air, and they persist in the environment and are transmitted to new hosts (Barbut, Gariazzo et al. 2007).

C. difficile is an important nosocomial pathogen in part due to the healthcare facility environment containing a high number of spores from infected patients; the hospital environment and patients taking antibiotics develop a discrete ecosystem where *C. difficile* persists and where certain virulent clones can survive. A major risk factor for CDI is age (≥65 years); this is generally believed to be due to a senescence of the immune response,

resulting from a combination of comorbidities, immune-related changes in the fecal flora and normal age-related changes (Ginaldi, Loreto et al. 2001).

In addition, alterations of the intestinal microflora allow the overgrowth of *C. difficile* caused by the consumption of broad-spectrum antibiotics, such as clindamycin, penicillins, cephalosporin and the flouroquinolones (Schroeder 2005; Bartlett 2008; Hookman and Barkin 2009).

Other contributory factors for CDI include: gastrointestinal surgery; chemotherapeutic agents for cancer, for example, methotrexate and proton-pump inhibitors, all of which affect the gastrointestinal microflora allowing *C. difficile* proliferation followed by the production of high levels of toxin(s). Several virulence factors have been described in *C. difficile* such as flagella (Eveillard, Fourel et al. 1993), fimbriae and proteolytic enzymes (Borriello, Davies et al. 1990), surface layer proteins (Bianco, Fedele et al. 2011). Among these, toxins A and B are recognized as the major factors responsible for CDI (Voth and Ballard 2005).

One of the preconditions for the beginning of the infection is the colonization by *C. difficile* of the host's intestinal tract and evasion of the immune system. The organism will then enter the host and survive numerous pressures to compete with the flora of the gastrointestinal tissue, produce and secrete the major virulence factors: toxins tcdA and tcdB (Wright, Drudy et al. 2008)

C. difficile produces three toxins: TcdA, TcdB and the binary toxin (CTD). Several studies have reported that the most important virulence factors are the two toxins A and B, showing that the main symptoms of infection are due only to their action (Rupnik, Wilcox et al. 2009).

The genes encoding them, *tcdA* and *tcdB*, are chromosomally located along with three accessory genes forming the 19.6-kb pathogenicity locus (Figure 1). These are: *tdcR*, which encodes an alternative RNA polymerase sigma factor positively regulating toxin production; *tcdC*, a negative regulator of toxin production that interferes with the RNA polymerase-TcdR complex; and *tcdE*, a gene encoding a holin-like protein, involved in the release of toxins (Dupuy, Govind et al. 2008).

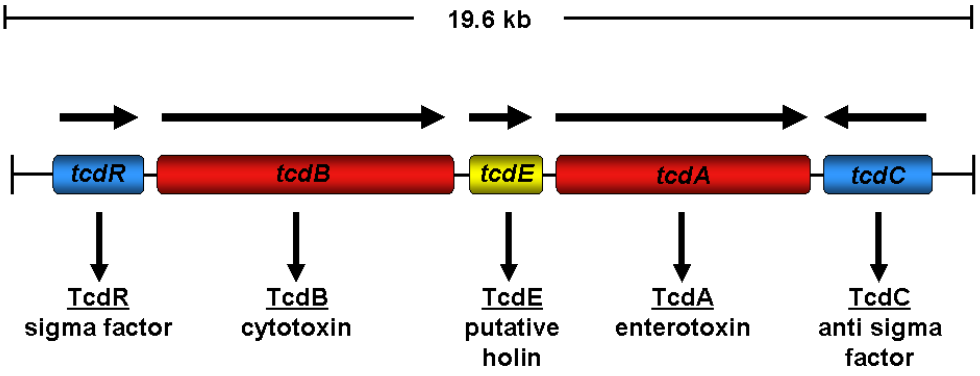


Figure 1 Pathogenicity Locus (PaLoc) of *Clostridium difficile* 630 encodes the two large clostridial toxins TcdA and TcdB and the two regulatory proteins TcdR and TcdC

Both toxins belong to the family of large clostridial toxins (LCT) characterized by a size between 260-308 kDa, by cytotoxic activity and by a common mechanism of action (von Eichel-Streiber, Boquet et al. 1996).

After their release into the environment, they enter in eukaryotic cells where they act with a mono-O-glucosylation (of a specific threonine residue) on Rho / Ras proteins, that are essential for many molecular mechanisms. As all the glucosyltransferase toxins, toxin A and toxin B are organized into four domains "ABCD":

A: biological activity domain

B: receptor binding domain

C: auto proteolitic cleavage domain

D: delivery domain

The domain A has glucosyltransferase activity, is at the N-terminus of the toxin and is the only domain of the toxin to be translocated into the cytosol of the target cell. The B domain, at the C terminus, is involved in binding with receptors rich in carbohydrates and is characterized by regions composed of repeats. The domain C, situated directly downstream of the glucosyltransferase domain, has a self-protease activity and is responsible for the processing of the toxin. The domain D is the least characterized and is thought to be involved in the formation of transmembrane structures, during the formation of pores and in the translocation of the toxin into the cytosol (Belyi and Aktories 2010) (Figure 2).

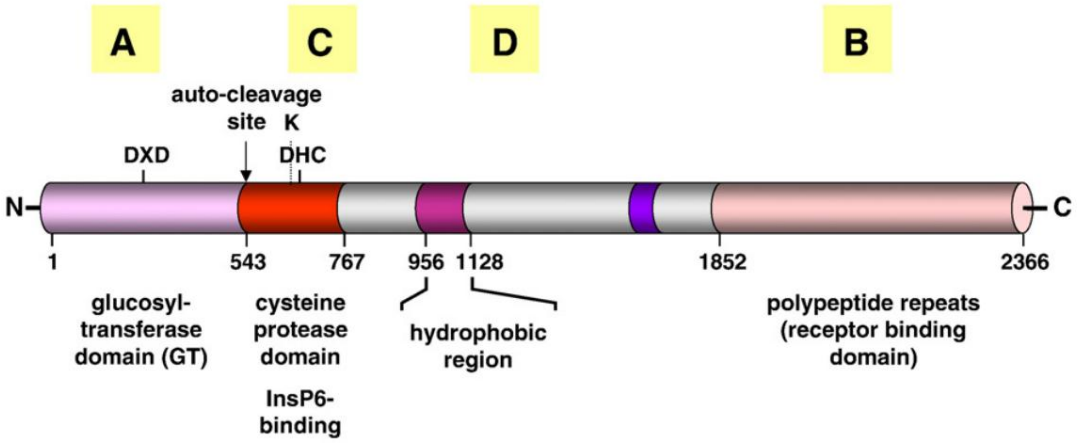


Figure 2. ABCD-model of clostridial glucosylating toxins. The clostridial glucosylating toxins are constructed of at least 4 domains. The A-domain covers the glucosyltransferase activity. The B-domain consisting of polypeptide repeats is involved in receptor-binding. The C-domain is responsible for the autocatalytic cleavage of the toxins (arrow: cleavage site) and is a cysteine-protease with the catalytic residues DHC. Lysine-600 (K) was shown to be involved in InsP6-binding. InsP6 is necessary for activation of the cysteine protease. The D-domain is likely involved in the delivery of the A-domain into the cytosol. This domain contains a hydrophobic region (indicated) suggested to be important for insertion of the toxin into endosome membranes.

The two toxins have different tropisms for the host cell membrane: tcdA binds the apical portion of the cell, while tcdB prefers to attach to its basolateral portion (Rupnik, Wilcox et al. 2009) (Figure 3).

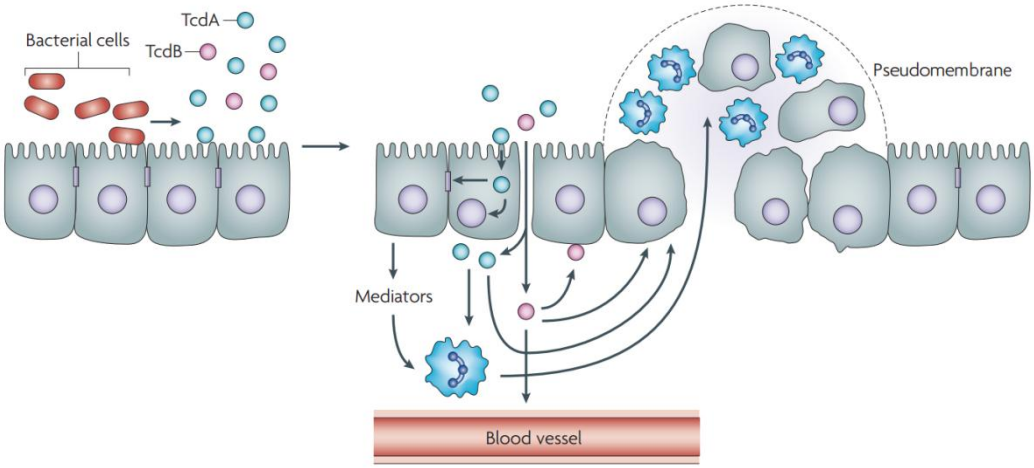


Figure 3. *C. difficile* colonizes the intestine (colon) after disruption of the normal intestinal flora. To what extent adhesion and biofilm production are involved in the pathogenesis of *C. difficile* is unknown; in the scheme, bacterial cells are shown as free cells and attached to host cells. Toxigenic strains produce toxin A and toxin B (TcdA and TcdB). TcdA binds to the apical side of the cell and, after internalization, causes cytoskeletal changes that result in disruption of tight junctions and loosening of the epithelial barrier, in cell death or in the production of inflammatory mediators that attract neutrophils. Disruption of tight junctions enables both TcdA and TcdB to cross the epithelium. TcdB binds preferentially to the basolateral cell membrane. Both toxins are cytotoxic and induce the release of various immunomodulatory mediators from epithelial cells, phagocytes and mast cells, resulting in inflammation and the accumulation of neutrophils. In an animal model, TcdB was shown to have a tropism for cardiac tissue, which would require that TcdB enters the bloodstream.

After endocytosis, the proteolytic activity of TcdA and TcdB leads to cleavage of the catalytic domain from the holotoxin, which is then transferred into the cytoplasm through a toxin-mediated pore. This cleavage requires only inositol phosphate from the host cell as a co-substrate (Rupnik, Wilcox et al. 2009).

In the cytosol the two toxins glucosylate the small GTPase Rho and Ras superfamily that are involved in many cellular processes: organization of the cytoskeleton, cell motility, regulation of transcription, cell cycle progression and apoptosis (Belyi and Aktories 2010).

Besides the damage caused by TcdA and TcdB, there could be additional damage caused by the action of the binary toxin CDT (Popoff, Rubin et al. 1988; Goncalves, Decre et al. 2004). The role of this protein in pathogenesis is not yet defined, but it is known to act inside the host cell by modifying the polymerization of actin through ADP-ribosylation (Schwan, Stecher et al. 2009; Papatheodorou, Wilczek et al. 2012).

CDI can be treated with a limited number of agents such as metronidazole and oral vancomycin (Owens 2007). Newer agents such as fidaxomicin have recently been licensed, though more knowledge of their effects is stiil needed from clinical practice (Louie, Miller et al. 2011).

The use of fecal transplantation from healthy people to patients suffering CDI is an additional method which has been practiced in refractory cases of CDI with the aim of restoring normal microbiota. Furthermore, vaccination may lead to an improvement of the current treatment options (Lo Vecchio and Zacur 2012).

A variety of molecular typing approaches have been developed to study *C. difficile* and CDI including: multilocus sequence typing (MLST), multilocus variable-number tandem-repeat analysis (MLVA), *slpA* gene sequence typing, amplified fragment length polymorphism, pulsed-field gel electrophoresis (PFGE), restriction endonuclease analysis (REA), toxinotyping (based on sequence data of toxins A and B) and PCR ribotyping (Killgore, Thompson et al. 2008). Generally, most of the methods are compatible, though PCR ribotyping is the most widely accepted in Europe with the Anaerobe Reference

Laboratory in Cardiff (UK) having a collection of the strains and also the role of assigning PCR ribotypes; more than 427 PCR ribotypes have been identified. In North America, PFGE is the preferred method of typing.

The PCR ribotyping is a typing method based on the amplification of the intergenic spacer region (ITS) between 16s and 23s rDNA of *C. difficile*. The number and length of ITSs vary in the genome of different *C. difficile* strains. Consequently, different migrations patterns of migration of the PCR products are observed for different strains. Each pattern of migration corresponds to one PCR ribotype (Figure 4).

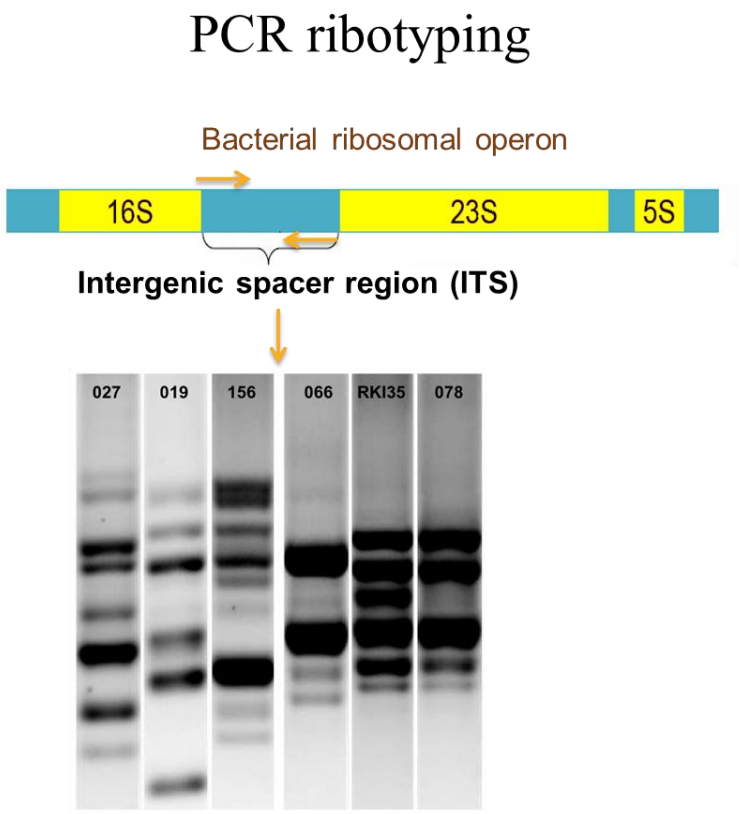


Figure 4. Schematic view of the PCR ribotyping method.

Alarmingly, in the past decade the emergence of a new group of highly virulent *C. difficile* strains (with PCR ribotype 027) has been reported that has caused outbreaks of increased disease severity in North America and Europe. In 2003 an epidemic of CDI was reported in southern Quebec and the Montreal district of Canada (Pepin, Valiquette et al. 2004).

Concurrently, a similar strain had also been isolated in several states of North America, and in March 2004, Stoke Mandeville Hospital (in the UK) reported a major outbreak including severe cases of CDI: 334 cases and 38 deaths (Cairns, Stabler et al. 2012). By 2005,

a considerable number of hospitals were reporting rates of CDI at five times their baseline rate (Loo, Poirier et al. 2005). Outbreaks in the UK, North America and Canada were found to be due to the dissemination of an epidemic strain subsequently identified as BI by REA, NAP1 by PFGE and PCR ribotype 027 by PCR ribotyping and referred to as BI/NAP1/027 (Killgore, Thompson et al. 2008). In late 2005, other European countries documented outbreaks of this strain (van Steenbergen, Debast et al. 2005; Kuijper, Barbut et al. 2008; Bacci, St-Martin et al. 2009) and similar reports from many other countries across the globe including Australia (Riley, Thean et al. 2009), Korea (Tae, Jung et al. 2009), Japan (Kato, Ito et al. 2007), Hong Kong (Cheng, Yam et al. 2009) and Costa Rica (Quesada-Gomez, Rodriguez et al. 2010) were subsequently announced. Patients infected with PCR ribotype 027 were found to have more severe diarrhea, higher mortality and more re-occurrences of symptoms (Borriello 1998). It has been thought that this increase in virulence and increase in re-occurrence rate were due to one or more of the following characteristics: exposure to fluoroquinolone antibiotics prior to CDI; increased production of toxin; prolonged production of toxin; and increased sporulation, which in turn increases the risk of transmission. Exposure to the fluoroquinolone class of antibiotics has been thought to be the strongest risk factor for CDI with PCR ribotype 027 (McCusker, Harris et al. 2003).

Epidemic strains of PCR ribotype 027 have been reported to produce significantly higher levels of toxins A and B compared with other strains (McDonald, Killgore et al. 2005). PCR ribotype 027 was also found to have an 18-bp deletion and a frameshift mutation due to a single base pair deletion at position 117 in the *tcdC* gene (Dupuy, Govind et al. 2008). The frameshift mutation results in a truncated protein and it is hypothesized that this leads to the deregulated expression of toxins A and B.

Rates of CDI, most notably PCR ribotype 027 in the UK and other parts of Europe, appear to have declined in the last 4 years. Furthermore, between 2007/2008 and 2010/2011, there was a 42.9% decrease in the number of PCR ribotype 027 strains isolated by the *C. difficile* Ribotype Network (CDRN) in the UK (Cairns, Stabler et al. 2012). This has occurred simultaneously with an increase in a variety of other PCR ribotypes, and now PCR ribotypes 014, 001 and 078 represent the most prevalent strains in European hospitals, as observed in the last European surveillance performed in 2008 (Bauer, Notermans et al. 2011). Emergence of the animal-associated PCR ribotype 078 *C. difficile* is recognized as both a gut colonizer and a cause of CDI in domestic animals and livestock (Cairns, Stabler et al. 2012). PCR ribotype 078 isolates have been found to be the most predominant from animal species with CDI, most notably pigs, calves and horses (Rupnik, Widmer et al. 2008). *C. difficile* has also been found in contaminated food for human consumption, with PCR ribotype 078 being the

most frequently implicated (Simango and Mwakurudza 2008; Rodriguez-Palacios, Reid-Smith et al. 2009; Songer, Trinh et al. 2009).

Recently, a study found *C. difficile* in 4.8% (five out of 119) of seafood and fish samples from a grocery store and all toxin positive isolates were found to be PCR ribotype 078 (Metcalf, Avery et al. 2011).

A population-based cohort study found that community-acquired CDI affected those who lack the traditional risk factors such as hospitalization or antibiotic exposure, and patients with community-acquired CDI were also found to be younger (Khanna, Pardi et al. 2012). The documented cases of community-acquired CDI also appear to be on the rise, and the increased incidence of community acquired infection with PCR ribotype 078 could be linked to the fact that this strain is found in both humans and animals (Cairns, Stabler et al. 2012). In addition, with the increase in documented cases of isolation of *C. difficile* from food products for human consumption, concerns have been raised about possible transmission between animals and humans in the community. Although *C. difficile* is not a proven food borne pathogen, there is evidence that the same strain can cause symptomatic disease in both pigs and humans (Debast, van Leengoed et al. 2009).

As PCR ribotype 027, PCR ribotype 078 has been described to be an emerging pathogen.

The characterisation of the genetic conservation between strains belonging to different PCR ribotypes could help significantly to identify the genetic elements associated with the onset of CDI. It has already been established through sequencing of the genomes of a number of *C. difficile* strains of different origin that there is a high variability in gene composition and conservation between strains (Sebaihia, Wren et al. 2006; He, Sebaihia et al. 2010; Scaria, Ponnala et al. 2010). In this study, we analyse the distribution and variability of 14 *cwp* genes in 40 *C. difficile* clinical isolates of the 6 prevailing PCR ribotypes in Italy (Spigaglia, Barbanti et al. 2010) and, more generally, in Europe (Bauer *et al.*, 2011).

1.2 The "Reverse Vaccinology" approach

Vaccines can be made from live-attenuated microbes, inactivated microorganisms, and purified microbial components such as polysaccharide-carrier protein conjugates. These approaches lead to the development of a conventional vaccine. However, they show several limitations. For example, immunogenic proteins are not necessarily protective antigens, often have a variable sequence and they are difficult to produce and/or purify on a large scale, leading to high production costs. In addition, the antigens chosen are only the most abundant

and not all of them can be analysed simultaneously. As a consequence these traditional approaches have failed against several infectious diseases and, thus, vaccines have not yet been developed.

Bioinformatics and whole-genome sequencing of bacteria brought innovation to the vaccinology field, identifying potential vaccine candidates without the need to grow the pathogen. This approach, named 'reverse vaccinology' (Rappuoli 2000), reduces the time and cost of new vaccine identification also for those diseases for which conventional approaches have failed (Mora, Veggi et al. 2003). Scientists can now search for the potential surface protein sequences using various algorithms. This new approach allows systematic identification of all the potential antigens of a pathogen in order to develop a safe and efficacious vaccine against any infectious disease.

After the genes coding for the pathogen proteins have been identified, they are cloned in *E. coli*, expressed and the corresponding proteins purified. Purified proteins are then used to immunize mice and the sera analysed in order to verify their predicted exposure and ability to trigger an immune response. Protective antigens are analysed again *in vitro* using opsonophagocytosis and bactericidal assays with the aim of final antigen selection. Potential candidates can also be determined on the basis of sequence conservation among different strains of the pathogen (Figure 5).

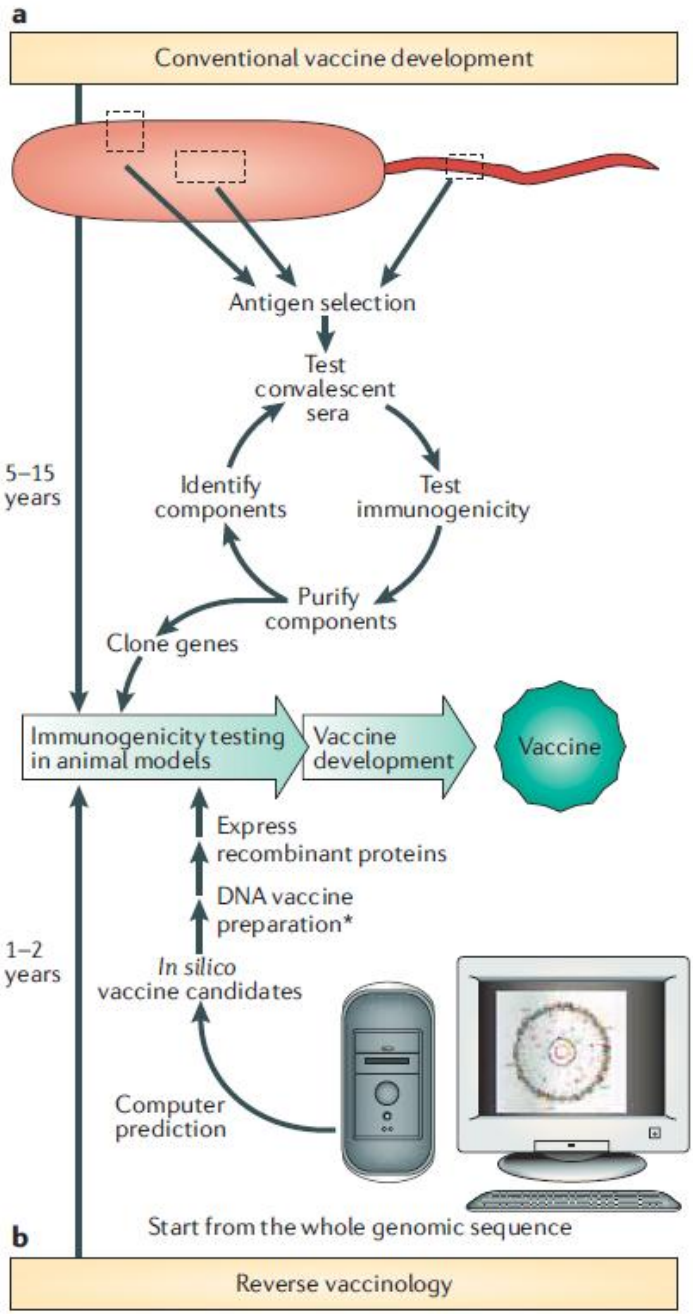


Figure 5. Approaches to vaccine development. Schematic demonstration of the essential steps required for vaccine development using the conventional approach (a), and reverse vaccinology (b). (Johri, Paoletti et al. 2006)

The sequencing of a bacterial genome gives the possibility to discover novel antigens missed by conventional vaccinology methodologies. The reverse vaccinology approach was first used to identify antigens for the development of a vaccine against serogroup B meningococcus (Mora, Veggi et al. 2003). The use of multi-genome sequence information for vaccine design represented a major change from the common concept that a single genome sequence is sufficient to produce a potential vaccine candidate. A single genomic sequence is not sufficient to represent the variability of bacterial populations. Therefore,

multiple sequences might be needed for a vaccine formulation that is effective in the case of a highly differentiated species. This is common to many important bacterial pathogens and this observation is in the context of the pan-genome reverse vaccinology era (Mora, Donati et al. 2006). Thus a universal vaccine can be obtained by using a combination of antigens chosen from different strains and not only a single one. However, their selection should consider the population structure of the microorganism, giving importance to each representative strain in the epidemiology of the disease.

The next step in order to reach a more comprehensive picture of bacterial populations will be population vaccinology. Vaccines will be then formulated from a collection of proteins that, together, protect against the major circulating populations of a pathogen. In addition, the sequencing of human and pathogen genomes has provided large amounts of data relevant to the study of human immune responses and complex host–pathogen interactions. Using and ameliorating immuno-informatic tools, such as T-cell and B-cell epitope-mapping algorithms, and of structure-oriented bioinformatics (Arcus, Lott et al. 2006; De Groot 2006) will lead to the refinement of the totally synthetic vaccine design containing strings of the best epitopes encoded by the microorganism (Figure 6).

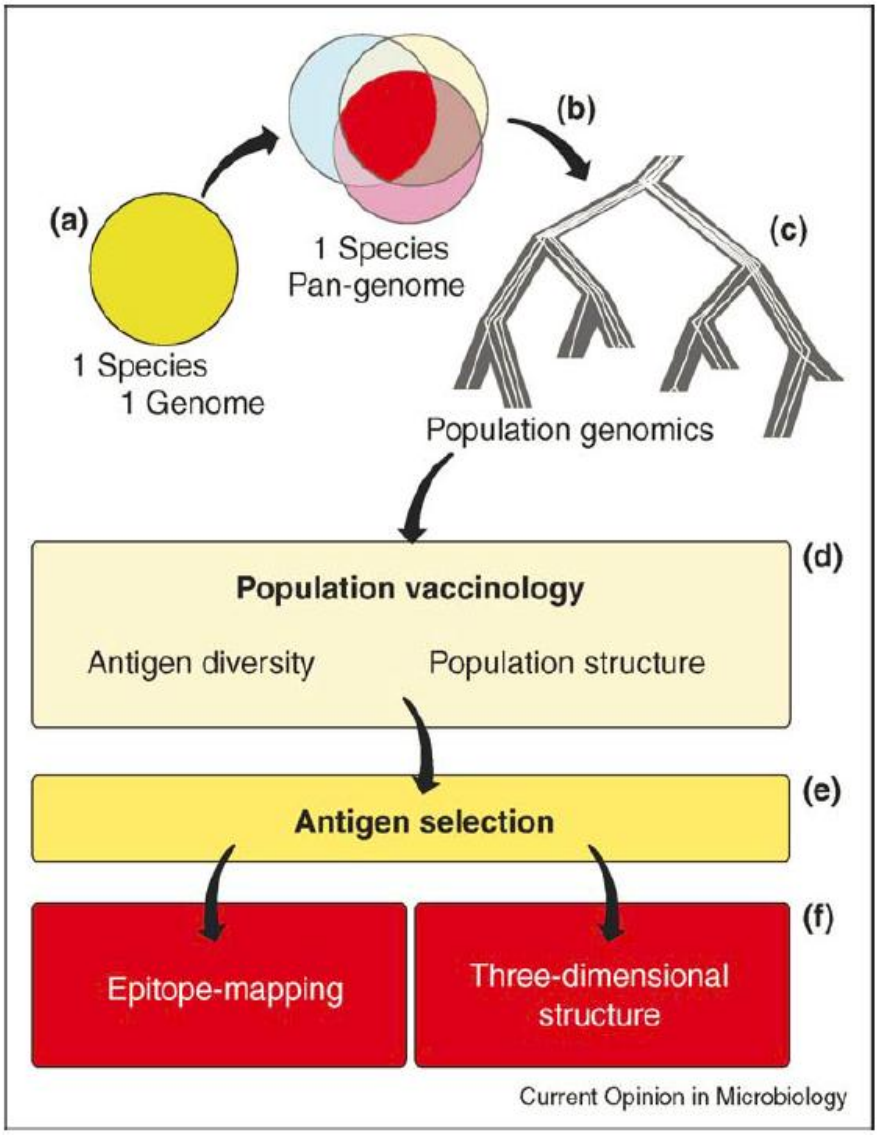


Figure 6 Flow chart for antigen discovery and refinement of the search. Three major genome-based approaches are involved in the identification of new potential vaccine candidates: (a) analysis of a single genome sequence in order to select secreted or extra-cellular proteins to identify potential vaccine candidates, (b) comparison of multiple genomes of the same species to assess intra-species diversity, (c) population genomics to achieve a more comprehensive coverage against the major circulating species. These three steps lead to (d) population vaccinology, which takes into account antigen variability and population structure, (e) allowing a more rational design of a new generation of vaccine targets, and (f) in silico screening, such as epitope-mapping and structure-oriented bioinformatics, will enble refinement of the search (Mora, Donati et al. 2006).

In the case of *C. difficile*, 300 proteins have been identified as potential antigens by a reverse vaccinology approach based on sequence analysis of the genome of 10 different strains. Among these 300 putative antigens, a major group of proteins belonging to the same family is represented by the cell wall binding proteins or CWPs.

In this study, we analyse the distribution and variability of *cwp* genes using two approaches. We have analysed the distribution of 14 cwp genes in 40 clinical isolates and

their conservation with respect to strain 630, we then focused on the characterisation of the *in vitro* expression of Cwp proteins (by Western blots) and also on the analysis of expression of Cwp proteins *in vivo* using a protein microarray approach.

1.3 The cell wall proteins

One of the main components of the bacterial surface is the S-layer, a protein structure that forms a regular two-dimensional network on the whole surface of the bacterium and is present in many prokaryotes (Calabi, Ward et al. 2001). The S-layer is involved in virulence also in the case of other human pathogens (McCoubrey and Poxton 2001; Eidhin, Ryan et al. 2006). However, although the proteins that consitute the S-layer belong to the same protein family, between different species such proteins have a low or no similarity of sequence (Fagan, Albesa-Jove et al. 2009).

The S-layer completely covers the surface of the bacterium and, acting as an adhesin, facilitates the bacteria-host cell interaction (Fagan, Albesa-Jove et al. 2009).

The S-layer described for the first time in *C. difficile* by Kawata *et al*., in 1984, consists of two proteins of different molecular weight, which associate to form a solid complex (Fagan, Albesa-Jove et al. 2009). The molecular weight of the two subunits varies depending on the strain: the larger of the two proteins may vary from 48 to 56 kDa, whereas the smaller varies from 37 to 45 kDa (McCoubrey and Poxton 2001; Eidhin, Ryan et al. 2006).

These proteins are encoded by the *slp*A gene as a single precursor. After a post translational cleavage of this precursor two mature proteins are produced, the HMW (high molecular weight) SLP and the LMW (low molecular weight) SLP (Figure 7). The SLPs facilitate adhesion to cultured cell lines and the LMW SLP is an immunodominant antigen (Calabi, Ward et al. 2001).

Figure 7. The precursor protein SlpA showing the cleavage sites generating the signal peptide (▲) and the mature HMW SLP and LMW SLP (△).

This is the first reported example in which the two S-layer proteins are derived from a single gene product and not by the expression of two separate genes (Calabi, Ward et al. 2001). The precursor, responding to a signal sequence, is conducted through the cytoplasmic membrane to be processed into two proteins: the *high molecular weight* (HMW) SLP and the *low molecular weight* (LMW) SLP (Calabi and Fairweather 2002).

These proteins are generated by posttranslational cleavage of the precursor SlpA by the cysteine protease Cwp84 (de la Riva, Willing et al. 2011). The two subunits are positioned in such a way that the HMW, facing the interior, acts as an anchor for the LMW exposed to the external environment (Fagan, Albesa-Jove et al. 2009) (Figure 8).

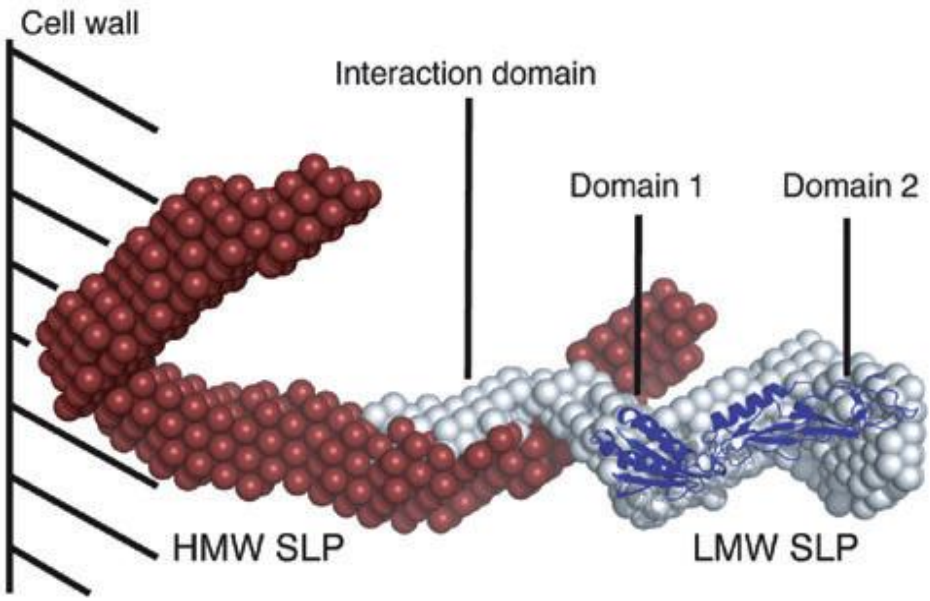


Figure 8. Model of orientation of the subunits of the S-layer. The HMW SLP is shown interacting with the cell wall through one region of the protein; however, the extent and exact nature of this interaction is currently unknown. The LMW SLP interacts, through the interaction domain, with the HMW and is directed towards the external environment (Fagan, Albesa-Jove et al. 2009).

After analysing several strains, the LMW-SLP shows high sequence variability (Fagan, Albesa-Jove et al. 2009). This finding can be explained by considering the position of this subunit. In fact, the exposure to the external environment and the marked variability of LMW-SLP represent a strategy to escape the host immune system (Fagan, Albesa-Jove et al. 2009).

Furthermore, a TBLAST analysis of the genome of strain 630 with the amino acid sequence of SlpA, revealed sequence homology between the HMW SLP and 28 other gene products, 11 located in the same locus of the *slpA* gene and 17 in other regions of the genome (Calabi et al. 2001). These 28 genes encode the cell wall proteins. The *slpA* locus of strain 630 is 36,661 bp long and includes 18 ORFs. Among these, only 12 ORFs, distributed both downstream and upstream of the *slpA* gene, code for Cwps (Calabi, Ward et al. 2001) (Figure 9).

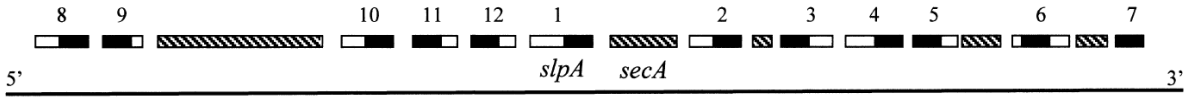


Figure 9. Arrangement of predicted ORFs upstream and downstream of the *C. difficile* 630 *slpA* gene. All ORFs are transcribed in the same direction (left to right), and the size of each protein is proportional to the size of the box. The regions of proteins shaded in black indicate where sequence homology is found with N-acetyl muramoyl-L-alanine amidase, and those not shaded where sequence homology is lacking. Proteins indicated by hatched shading are unrelated to N-acetyl muramoyl-L-alanine amidase.

Further analysis shows a significant sequence homology of HMW SLP with the family of autolysin of *Bacillus subtilis*, in particular with proteins LytCB and LytC. The latter is the main autolysin of *B. subtilis* and has an N-acetyl muramoyl-L-alanine amidase activity, involved in the destruction of the peptidoglycan. Although the involvement of these proteins in cell adhesion has been verified, the specific function has not yet been identified (Sebaihia, Wren et al. 2006).

Several of the *cwp* genes are not conserved in all the *C. difficile* genomes characterised so far. In general, 12 of the 29 *cwp* genes are clustered in the same region of the genome, named after *slpA* (*cwp1*) the *slpA* locus (Calabi, Ward et al. 2001; Karjalainen, Waligora-Dupriet et al. 2001), whereas the remaining 17 paralogs are distributed throughout the genome. We focused our study on the genes coding for the latter group of CWPs since they are poorly characterised.

Knowledge of the conservation of these genes in clinical isolates would offer useful information for the characterisation of the role that CWPs may have in *C.difficile* infection and also provide another tool for classifying newly emerging strains.

To date, 29 cwp genes have been identified in *C. difficile* which code for a family of cell wall proteins involved in colonisation and pathogenesis. All of these CWPs have a conserved domain containing two or three copies of the Pfam 04122 motif, a putative cell wall binding repeat 2 (Fagan, Janoir et al. 2011)(Figure 10).

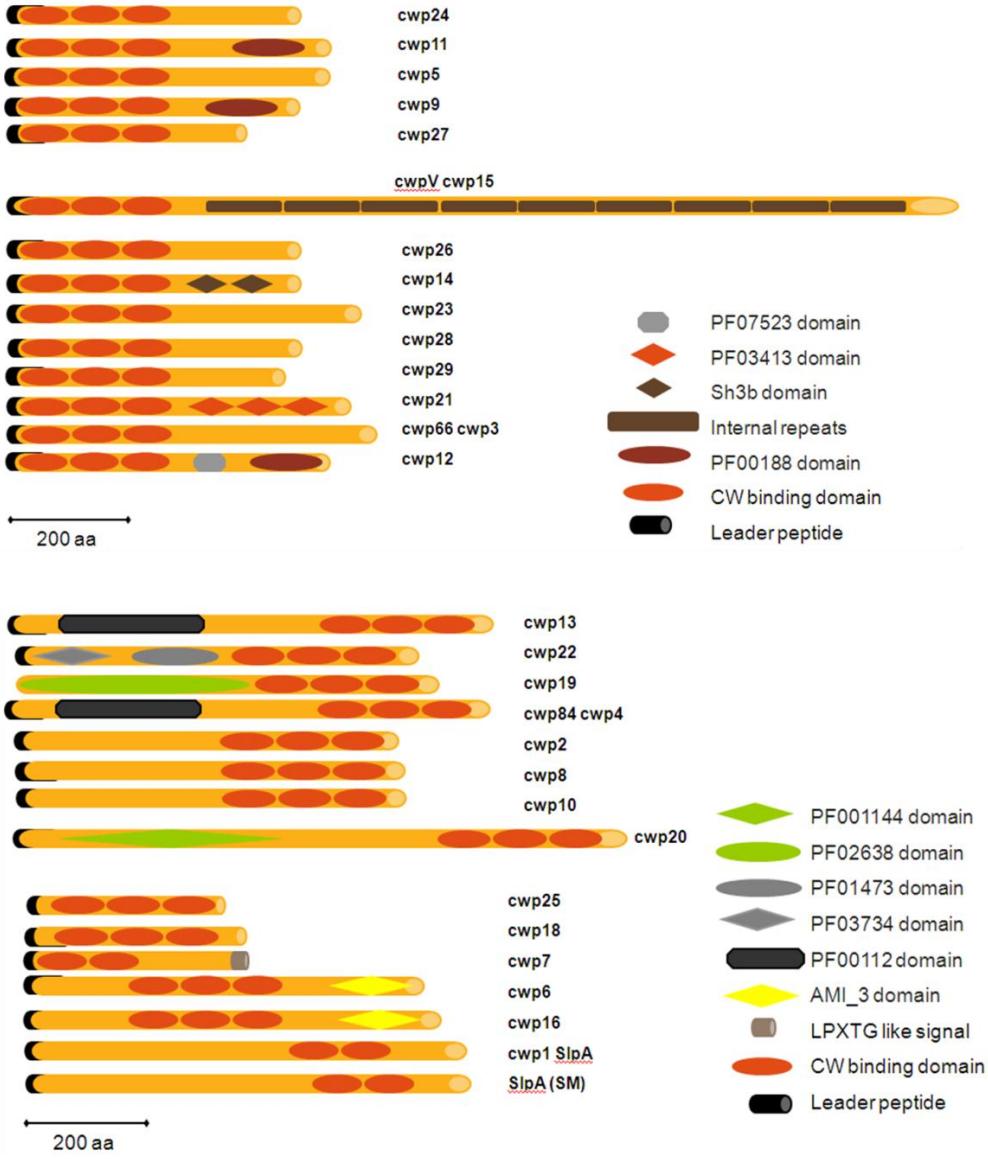


Figure 10. Schematic representation of the domain composition of the 29 CWPs of *C. difficile.*

In addition, several of the CWPs show a second, more variable domain that may specify a unique function. Fagan *et al.*, in 2009 suggested a model of the cell surface of *C. difficile* in which the continuous film of the S-layer is located above the peptidoglycan and is interrupted in several points by Cwps (Figure 11).

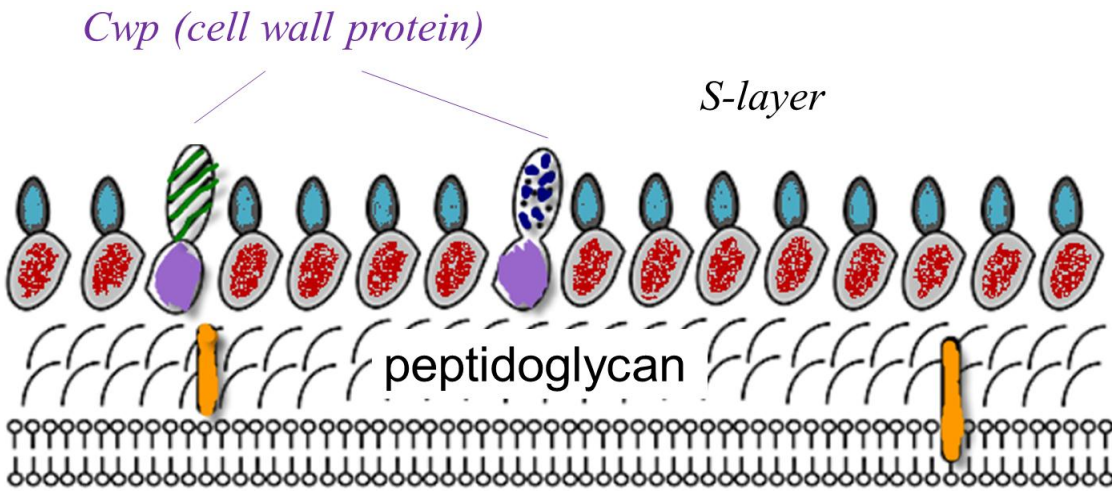


Figure 11. The two SLPs are shown above the peptidoglycan layer: the HMW SLP (red), the LMW SLP (blue). Other minor cell wall proteins are shown as two-lobed structures ; the filled areas, indicating domains predicted to be exposed to the environment, are variable between these proteins. Putative cell wall polymers (Ganeshapillai, Vinogradov et al. 2008) including putative lipid-containing polymers (Poxton and Cartmill 1982) are shown as vertical orange bars.

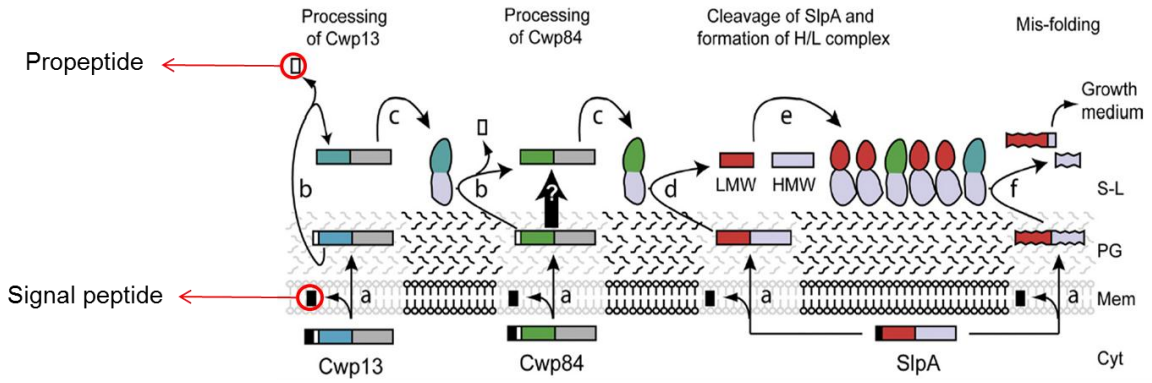
To better define the structure of the cell surface of *C. difficile*, the results of the studies by Davies and Borriello need to be taken into account. In 1990, they showed that many strains of *C. difficile* produced the capsule and that the genome of strain 630 contains a cluster of genes involved in the synthesis of extracellular polysaccharides (Davies and Borriello 1990; Sebaihia, Wren et al. 2006).

Current knowledge of the molecular basis of pathogenesis is limited primarily to the activities and regulation of two major toxins. In contrast, little is known about the mechanisms involved in colonization of the enteric system and the role of CWPs.

Some members of the CWP family have been investigated extensively. Among these, Cwp84 is a protease that cleaves the SlpA precursor and also degrades many proteins of the host cell extracellular matrix, Cwp66, which acts as an adhesin, and CwpV, a protein that is expressed in a phase variable manner (Emerson, Reynolds et al. 2009).

De la Riva *et al*. in 2011 compared the phenotypes of *C. difficile* strains containing insertional mutations in either *cwp84* or its paralog *cwp13*. They showed that the presence of

uncleaved SlpA in the cell wall of the *cwp84* mutant resulted in aberrant retention of other cell wall proteins at the cell surface. The *cwp13* mutant cleaved the SlpA precursor normally and had a wild-type-like colony phenotype. Both Cwp84 and Cwp13 are produced as proenzymes which are processed by cleavage to produce mature enzymes. In the case of Cwp84, this cleavage does not appear to be autocatalytic, whereas in Cwp13 autocatalysis was demonstrated (De la Riva, Willing et al. 2011). Cwp13 appears to have a role in the processing of Cwp84, but is not essential for Cwp84 activity. Cwp13 cleaves SlpA in the HMW SLP domain, and De la Riva suggested it may reflect a role in cleavage and degradation of misfolded proteins at the cell surface (Figure 12).



Lucía de la Riva et al. Journal of Bacteriology, July 2011

Figure 12 Model for processing and activities of Cwp84 and Cwp13. SlpA, Cwp84, and Cwp13 are produced as preproteins containing signal peptides that are removed during processing by the *sec* system (step a). The propeptides of Cwp84 and Cwp13 are removed (step b), either by autocatalysis in the case of Cwp13 or by an unknown activity together with Cwp13 activity in the case of Cwp84, to form the active enzyme species that are incorporated into the S-layer (step c). Mature Cwp84 cleaves the SlpA precursor (step d), which results in the formation of the H/L complex (step e). Misfolded proteins are recognized by Cwp13 and are cleaved in their cell wall binding domains to prevent incorporation into the S-layer, resulting in detachment from the cell and deposition into the growth medium (step f). S-L, S-layer; PG, peptidoglycan; Mem, membrane;Cyt, cytoplasm (de la Riva, Willing et al. 2011).

CwpV is the largest member of the CWPs family and is expressed in a phase variable manner (Reynolds, Emerson et al. 2011). Reynolds *et al.*, in 2011 showed that CwpV promotes *C. difficile* aggregation, mediated by the C-terminal repetitive domain. CwpV is post-translationally cleaved at a conserved site leading to formation of a complex of cleavage products (like the SlpA processing). The highly conserved N-terminus anchors the CwpV complex to the cell surface (Figure 13).

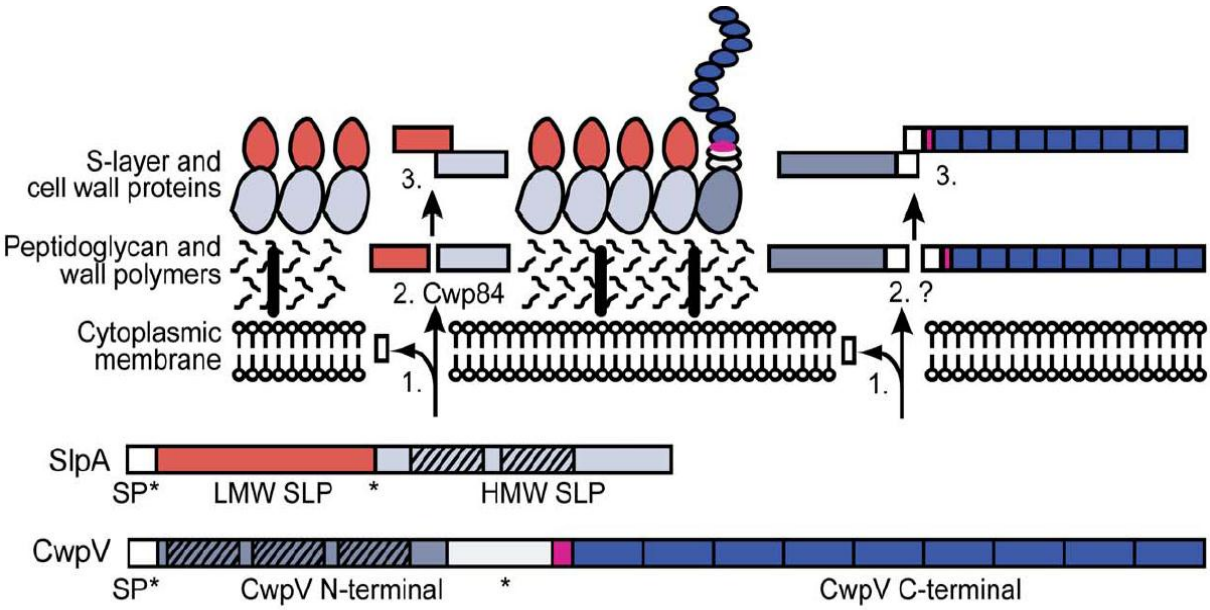


Figure 13.Cartoon representation of the overall model for post-translational CwpV processing and incorporation into the S-layer with analogy to SlpA processing. *,cleavage sites. Step 1, signal peptide (SP) cleavage and transport across the cell membrane. Step 2, cleavage of protein. SlpA is cleaved by Cwp84 into the LMW and HMW SLP. The protease responsible for CwpV cleavage is currently unknown. Step 3, formation of a complex of the products of cleavage, anchoring both products to the cell surface (Reynolds, Emerson et al. 2011).

Protein translocation across the cytoplasmic membrane is an essential process in all bacteria. The Sec system, comprising at its core an ATPase, SecA, and a membrane channel, SecYEG, is responsible for the majority of this protein transport (Driessen and Nouwen 2008). Recently, a second parallel Sec system has been described in a number of Gram-positive species (Bensing and Sullam 2002; Lenz and Portnoy 2002; Chen, Wu et al. 2004; Siboo, Chaffin et al. 2008). This accessory Sec system is characterized by the presence of a second copy of the energizing ATPase, SecA2, and, in the systems studied, is responsible for the translocation of a subset of Sec substrates.

In common with many pathogenic Gram-positive species, *C. difficile* possesses two copies of SecA. Fagan and Fairweather in 2011 described the first characterization of the *C. difficile* accessory Sec system and the identification of its major substrates, that are SlpA, Cwp2, CwpV, Cwp66 and Cwp84. Furthermore, they showed that expression of either dominant negative allele or antisense RNA knock-down of SecA1 or SecA2 dramatically impaired growth, indicating that both Sec systems are essential for *C. difficile* (Fagan and Fairweather 2011).

In this study 18 cell wall proteins have been selected for further characterization using a protein microarray analysis of the immune response obtained in a protection experiment carried out in hamsters. Briefly, hamsters were vaccinated with various domains of Toxin A (p5 /6) and Toxin B and then challenged with a lethal dose of *C. difficile* strains 630 or B1. Only animals that developed protective immunity would survive the lethal challenge. We have used the sera from protected and from control animals to identify antigens capable of eliciting a specific immune response towards *C. difficile* toxins and surface proteins.

Microarray technology has been a valuable approach to screen potential antigens useful for developing vaccines against *C. difficile*. Microarrays were implemented for the identification of the most immunogenic antigens and profiling of disease-specific antibody response to identify surface protein exposed during infection.

1.4 The Protein Microarray Technology

Traditionally, the properties of proteins have been elucidated by studying single molecules, one experiment at a time. Since this process is slow and labour intensive, in the last decade high-throughput methods have been developed in order to analyse a large number of molecules, such as DNA, proteins or metabolites simultaneously in a single experiment. In particular, DNA microarrays are a valuable tool in genomic research (Schena, Shalon et al. 1995). They have been used for several applications: gene expression patterns, location of transcription factor binding sites and detection of sequence mutations and deletions on a large

scale (Hall, Ptacek et al. 2007). However, DNA microarrays are only informative of the genes themselves and provide little knowledge regarding the protein functions they encode. More recently, the application of high-throughput approaches has been extended to protein studies. These include profiling with mass spectrometry (Gavin, Bosche et al. 2002; Ho, Gruhler et al. 2002; Gavin, Aloy et al. 2006) and protein microarrays (MacBeath and Schreiber 2000; Zhu, Bilgin et al. 2001).

Microarray technology allows the simultaneous analysis of many samples within a single experiment. The whole field of protein microarray technology has received considerable impetus as a result of the increasing genomic information available. New technologies such as automated protein expression and purification systems, used for the generation of capture molecules, and the need for analysis of whole 'proteomes' will be a driving force for fast developments within the field of protein microarray technology.

Protein chips have emerged as a new approach for a variety of applications including the identification of protein-protein interactions, protein-phospholipid interactions, small molecule targets, and substrates of protein kinases or of other enzyme families (Winssinger, Ficarro et al. 2002).

They can also be used for clinical diagnostics and to monitor diseases (Hall, Ptacek et al. 2007). Typically, protein chips are prepared by immobilizing proteins on a treated microscope slide using a contact spotter (MacBeath and Schreiber 2000; Zhu, Bilgin et al. 2001) or a non-contact microarrayer that applies capillaries or inkjet technology to deposit nanolitre–picolitre droplets onto the surface of the slide (Delehanty 2004; Jones, Gordus et al. 2006).

Proteins must remain in a wet environment. For this reason, the printing process is carried out in a humidity-controlled environment and the proteins are exposed to samples containing the corresponding binding molecules in solution. Different slide surfaces can be used for protein chips. When choosing a slide surface, the proteins should be immobilized on the chip maintaining their conformation and functionality, thus achieving maximum binding capacity (Zhu, Bilgin et al. 2003). Proteins can be oriented either randomly or uniformly on the slide surface. They can be attached randomly through amines, aldehyde- and epoxy-derivatized glass surfaces (Kusnezow, Jacob et al. 2003). Coating the glass surface with nitrocellulose, gel pads, or poly-L-lysine also leads to a random orientation as the proteins are passively adsorbed onto the surface (Angenendt, Glokler et al. 2002; Charles, Goldman et al. 2004). Reactive proteins can be located on a proteome chip with small molecule probes coupled to either fluorescent, affinity, photochemical, or radioisotope tags. Fluorescent labels are generally preferred as they are safe, effective and are compatible with readily available

microarray laser scanners. Antibodies are the most prominent capture molecules used to identify targets. Selectivity of the capture molecules is the most crucial issue in the context of all array-based proteomic approaches.

Protein chips are used in a wide range of applications; they have been used to unravel the functions of previously uncharacterized proteins and to discover new functions for known proteins. Proteome chips have been used to study protein-protein interactions (Zhu, Bilgin et al. 2001), protein-DNA interactions (Hall, Zhu et al. 2004), protein-lipid interactions (Zhu, Bilgin et al. 2001), protein-drug interactions (Huang, Zhu et al. 2004), protein-receptor interactions (Jones, Gordus et al. 2006), and antigen-antibody interactions (Michaud, Salcius et al. 2003). These microarrays could be used to study protein–protein interactions on a genome-wide scale, many known interactions could be confirmed and a set of novel binding proteins could be detected (Figure 14). Experiments designed to detect protein–lipid interactions have shown that the identification of proteins able to bind low molecular weight compounds is possible. This opens the possibility to examine an entire proteome directly for protein–drug interactions. Microarray immunoassays are of general interest also for all diagnostic applications where several parameters of one sample have to be analysed in parallel (Mendoza, McQuary et al. 1999; Schweitzer, Wiltshire et al. 2000).

In addition, proteome chips have been used to study kinase activities (Ptacek, Devgan et al. 2005) and serum profiling (Zhu, Hu et al. 2006). Proteomic research and diagnostic applications will be the two major fields addressed by protein microarray technologies. In medical research, protein microarrays will accelerate immune diagnostics significantly by analysing in parallel all relevant diagnostic parameters of interest. The reduction of sample volume is of great importance for all applications in which only minimal amounts of samples are available. One example might be the analysis of multiple tumour markers from a minimum amount of biopsy material.

Proteome chips have also been used successfully to screen patient's sera for the presence of autoantibodies (Kattah, Alemi et al. 2006) or viral specific antibodies (Zhu et al., 2006). Thus proteome chip technology is a valuable high-throughput method for probing a group of proteins for a specific function or property. It is an exceptional new way to discover unknown proteins, as well as new functions of already known proteins.

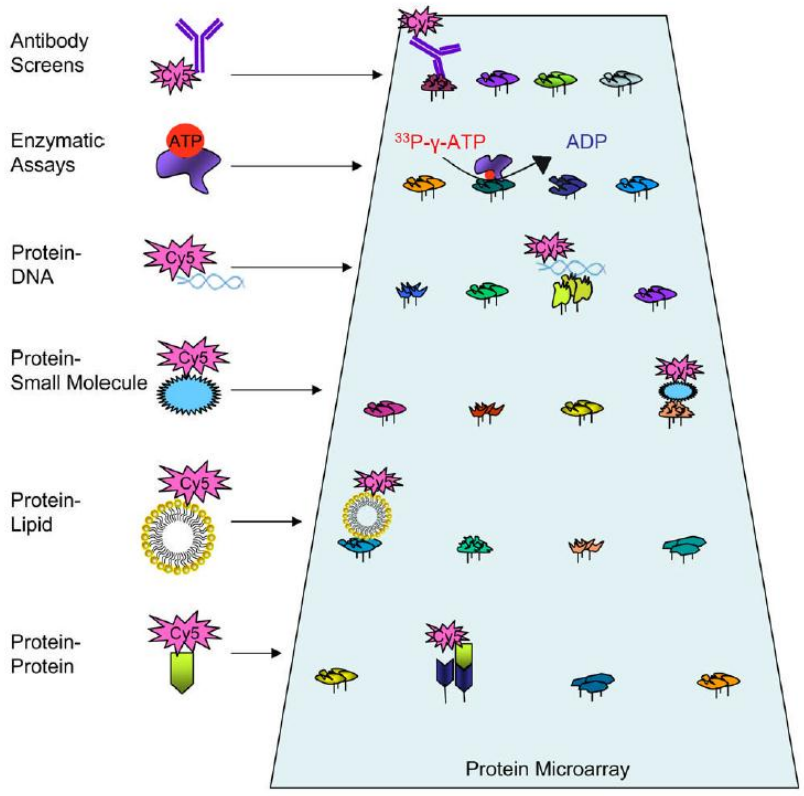


Figure 14 Applications of functional protein microarrays. A representative sample of the different assays that have been performed on functional protein microarrays. Proteins are immobilized at high spatial density on a microscope slide and the slide can then be probed for various interactions. While Cy5 is the fluorophore shown, many other fluorophores can be used for detection (Hall, Ptacek et al. 2007).

The growing field of protein microarray technology also requires the development of methods for high-throughput generation of recombinant proteins. Such methods are a prerequisite for the growing demand for thousands of specific capture molecules. In addition to their use in the generation and isolation of appropriate capture molecules, recombinant proteins will be used to generate microarrays that allow a rapid and efficient screening for high-affinity binders with minimal, or no, cross-reactivity to other proteins.

Detection of captured targets is performed by laser scanners with confocal detection optics (Templin, Stoll et al. 2002). Readout systems based on different techniques such as fluorescence, chemiluminescence or radioactivity can be used to detect complex formation within each spot (Figure 15).

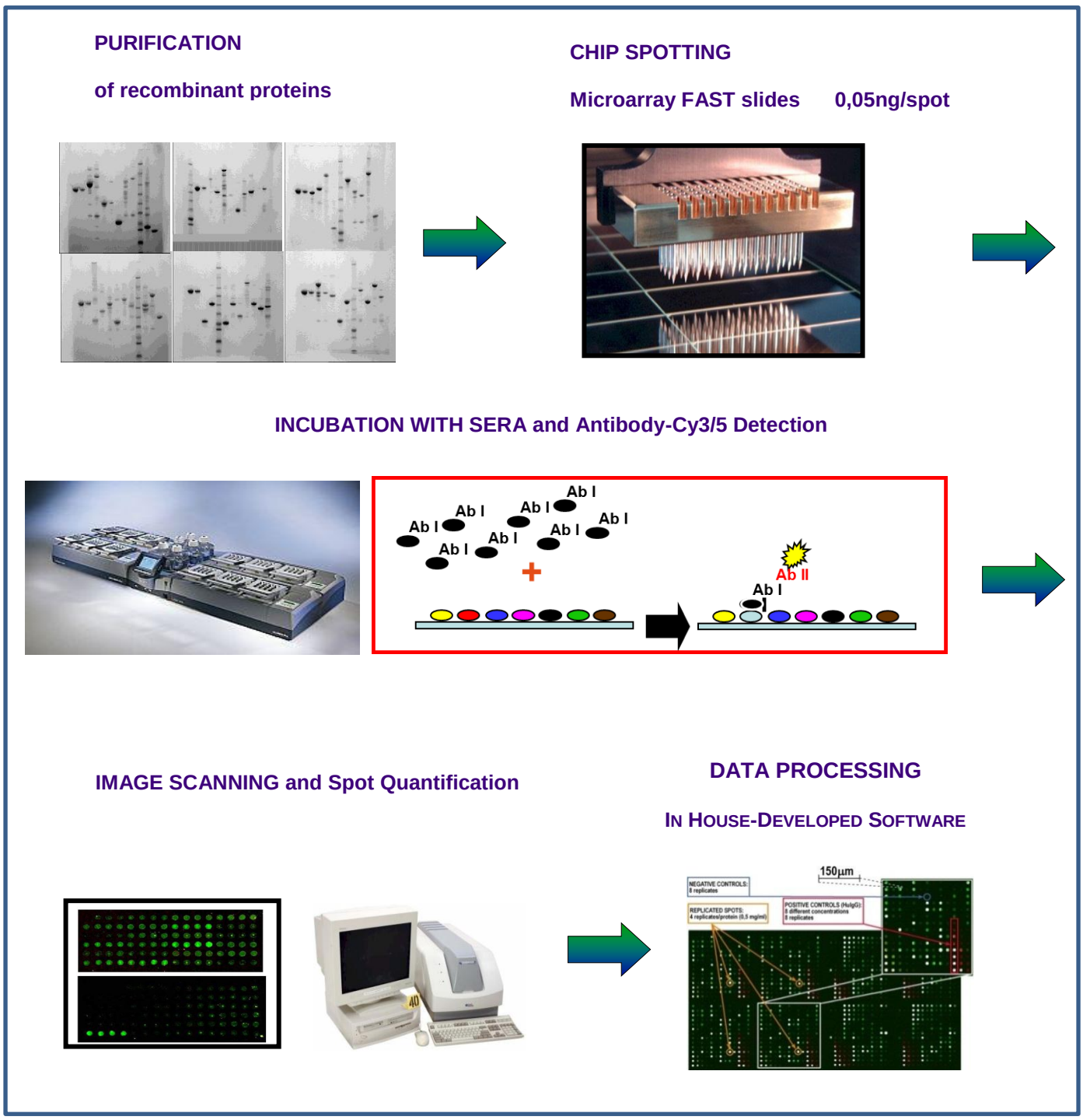


Figure 15. Schematisation of the protein microarray technique.

New trends in technology, mainly in nanotechnology and microfluidics, newly established detection systems and improvements in computer technology and bioinformatics have been rapidly integrated into the development of microarray-based assay systems. Microarrays built from tens of thousands of different probes per square centimetre, are now well established high-throughput hybridization systems that generate huge sets of data within a single experiment.

Accurate quantification with protein microarrays can be achieved by including positive and negative control spots and/or internal calibration spots. Hence, this will lead to robust and reliable diagnostic assays.

2. Materials and Methods

2.1 Bacterial strains and growth conditions

C. difficile clinical isolates collected by the Istituto Superiore di Sanità, Italy (1987-2010) were used in this study (Table 1). Strains were isolated from symptomatic patients in 13 different Italian hospitals. In particular, strains C192, C193, C252, C253, AR1, AR2, TR2, TR3, An45 and An56 were isolated during five different outbreaks that occurred in hospitals C, D, E, F and G. *C. difficile* isolates were typed as PCR ribotype 001 (two isolates), 012 (ten isolates), 014 (two isolates), 018 (ten isolates), 078 (ten isolates) and 126 (six isolates). Strains were cultured at 37°C under anaerobic conditions on *Brucella* agar plates containing vitamin K1 (0.5 mg/ml), haemin (5 mg/ml) and 5% defibrinated sheep red blood cells or in brain heart infusion (BHI) broth (Difco Laboratories). *C. difficile* strains 630, R20291 and M120 were used as control strains for PCR ribotype 012, 027 and 078/126, respectively. *C. difficile* 630 was obtained from the University of Paris-Sud, Chatenay Malabry, France. *C. difficile* strains R20291 and M120 were obtained from the Microbial Pathogenesis Laboratory, Pathogen Genomics, Wellcome Trust Sanger Institute, Hinxton, United Kingdom.

The *E. coli* strains were grown in LB or HTMC medium.

- LB (Luria Bertani broth)

Composition: 10 g/L Trypton, 5 g/L Yeast extract, 10 g/L NaCl pH 7.6; 100 μg/ml ampicillin

- HTCM (High Throughput Complex Medium)

Composition: 15 g/L Glycerol (or Glucose); 0.5 g/L $MgSO_4$ (2mM); 30 g/L yeast extract (Difco); 16 g/L K_2HPO_4; 6 g/L KH_2PO_4; 200 μg/ml ampicillin; pH 7.35

-BHI (Brain Heart Infusion Broth)

Composition: 17.5 g/L Brain Heart Infusion, 10 g/L Enzymatic Digest of Gelatin, 2 g/L Dextrose, 5 g/L NaCl, 2.5 g/L Na_2HPO_4 , pH 7.4 ± 0.2 at 25°C

-Brucella agar plates:

Composition: 39 g/L Columbia Agar Base (OXOID), 5 g/L Yeast extract (OXOID). 0.5 mg/ml vitamin K1, 5 mg/ml haemin and 5% defibrinated sheep red blood cells

Table 1. *Clostridium difficile* clinical isolates analysed in this study

Isolate	Year	Origin[§]	TcdA	TcdB	Binary toxin	Toxinotype	PCR ribotype
F III 10	2005	A	+	+	-	0	001
IT0843 EU	2008	B	+	+	-	0	001
C192	1987	C	+	+	-	0	012
C193	1987	C	+	+	-	0	012
C252	1987	D	+	+	-	0	012
C253	1987	D	+	+	-	0	012
AR1	1987	E	+	+	-	0	012
AR2	1987	E	+	+	-	0	012
TR2	1989	F	+	+	-	0	012
TR3	1989	F	+	+	-	0	012
An45	1989	G	+	+	-	0	012
An56	1989	G	+	+	-	0	012
F II 3	2005	H	+	+	-	0	014
IT0855 EU	2008	D	+	+	-	0	014
IT0603	2006	A	+	+	-	0	018
IT0707	2007	A	+	+	-	0	018
IT0807	2008	I	+	+	-	0	018
IT0808	2008	I	+	+	-	0	018
IT0825	2008	L	+	+	-	0	018
IT0829	2008	L	+	+	-	0	018
IT0839 EU	2008	B	+	+	-	0	018
IT0840 EU	2008	B	+	+	-	0	018
IT0926	2009	M	+	+	-	0	018
IT0929	2009	M	+	+	-	0	018
CD5	1998	H	+	+	+	5	078
IT0810	2008	I	+	+	+	5	078
IT0834	2008	I	+	+	+	5	078
IT0901	2009	N	+	+	+	5	078
IT0909	2009	O	+	+	+	5	078
IT0918	2009	I	+	+	+	5	078
IT0919	2009	I	+	+	+	5	078
IT0925	2009	M	+	+	+	5	078
IT0936	2009	M	+	+	+	5	078
IT1003	2010	D	+	+	+	5	078
1991	2001	H	+	+	+	5	126
2350	2002	H	+	+	+	5	126
3360	2006	H	+	+	+	5	126
IT0820	2008	L	+	+	+	5	126
IT0824	2008	L	+	+	+	5	126
IT0846 EU	2008	I	+	+	+	5	126

§ Strains were isolated at 13 Italian hospitals (arbitrarily denominated A to O).

2.2 Cloning and expression of *Clostridium difficile* recombinant proteins

C. difficile ORFs were PCR-amplified using chromosomal DNA of strain 630 as a template. PCR products were cloned into a modified pET15b^{+} expression vector (Novagen) using the PIPE cloning method (Figure 16) (Klock, Koesema et al. 2008). The PCR-amplified vector (vector PCR, V-PCR) and the PCR-amplified ORFs (insert PCR, I-PCR) were mixed and used to transform the *E. coli* HK100 strain (Klock, White et al. 2005).

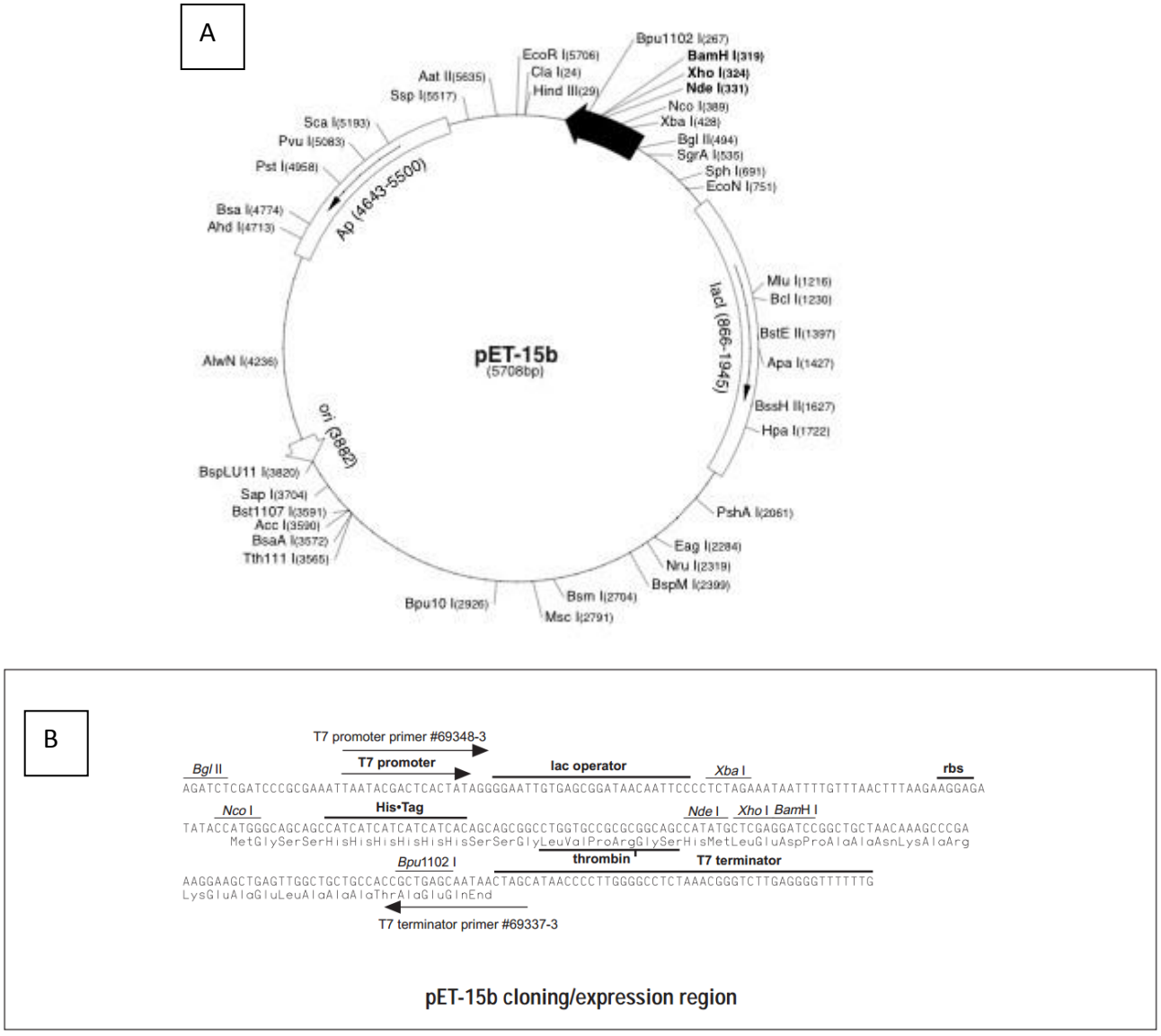


Figure 16. The pET-15b vector carries an N-terminal His-Tag sequence followed by a thrombin site and three cloning sites. Unique sites are shown on the circular map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown in panel B.

In order to reduce the number of transformants carrying the empty pET15b^{+} vector, used as V-PCR template, pET15b+ was modified by cloning the toxic CcdB gene into it using the PIPE technique. The CcdB-Chloramphenicol cassette was excised from the SpeedET vector (Klock, Koesema et al. 2008) and cloned into pET15b^{+} using the permissive *E. coli* strain DB3.1. After PCR colony screening and plasmid sequencing, a DB3.1 clone containing the correct pET15b-CcdB/cm construct was selected and used to amplify and purify the plasmid that was then used as DNA template for the V-PCR. The oligonucleotide primers used to prepare the V-PCR and the 15 bases at the 5' ends of the primers used to prepare the I-PCRs were designed in such a way that a Tobacco Etch Virus (TEV) cleavage site was introduced between the vector sequence coding for the N-term hexa-histidine tag and the ORF sequence. The forward and the reverse oligonucleotides used to obtain the V-PCR were 5'-TAACGCGACTTAATTCTAGCATAACCCCTTGGGGCCTCAAACGG-3' and 5'GCCCTGGAAGTACAGGTTTTCGTGATGATGATGATGATGGCTG CTGCCCATGGTATATC-3' respectively. The forward and reverse 15 base long 5' tails of the oligonucleotides used to obtain the I-PCRs were 5'-CTGTACTTCCAGGGC-3' and 5'-AATTAAGTCGCGTTA-3', respectively. 2 µl of V-PCR were mixed with 2 µl of each I-PCR and 3 µl of these mixtures were used to transform Ca^{2+} competent HK100 cells. Transformants were then selected by PCR screening and DNA sequencing of the plasmids extracted from the PCR-positive clones. In order to express cloned ORFs, the plasmids were prepared from the positive clones in HK100 and used to transform Ca^{2+}-competent BL21-DE3T1r (New England Biolabs) or the T7 Express competent *E. coli* cells (New England Biolabs). The new clones were screened for expression of the heterologous proteins. The wells of 96 deep well plates containing 1.5 ml of LB medium with the addition of 100 µg/ml ampicillin were inoculated with the expression clones and grown at 37 °C up to $OD_{600\ nm}$= 0.5. Protein expression was then induced by adding 1 mM IPTG and incubating the culture at the same temperature for an additional 3 h. To check protein expression and solubility, the induced cells were lysed with B-Per buffer (Pierce, Rockford, ILU.S.A.) using the manufacturer's instructions. The results were evaluated by analysing total and soluble protein extracts in SDS-PAGE.

2.3 *Clostridium difficile* DNA isolation, amplification and sequencing.

Genomic DNA was isolated by a standard protocol for Gram-positive bacteria using a NucleoBond AX-G kit (Macherey-Nagel) according to the manufacturer's instructions. Genes were amplified using primers specific for regions external to each ORF and are listed in Table 3. Only for *cwpV* were primers designed for amplifying a conserved internal segment (904 bp) of the otherwise highly variable coding region (Emerson, Reynolds et al. 2009). When primers were used in a multiplex PCR reaction, the two sets were added at different concentrations: 1 μM and 0.3 μM for primers specific for the longer gene and the shorter gene, respectively. DNA amplification was performed using 1 μl of purified genomic DNA (50 ng) in a final volume of 50 μl. The nucleotide sequences of PCR products were determined using a BigDye Terminator V3.1 kit (Applied Biosystem) in an ABI PRISM 3700 Analyzer (Applied Biosystems). The nucleotide sequences have been submitted to the GenBank database under accession numbers JQ389122-JQ389476.

2.4 Sequence alignments and phylogenetic analysis.

The percentage of sequence identity was calculated by pair wise BLAST with the VECTOR NTI SUITE 11 (Informax, Bethesda), with gaps included. Sequence alignments were performed using CLUSTAL W (1.83, GCG Wisconsin Package version 11.1) and phylogenetic trees were inferred by the neighbour-joining-distance-based method and bootstrapped 1,000 times.

2.5 Protein expression.

- ***C. difficile* protein extraction**

Total protein and S-layer extracts

The preparation of whole cell lysates was obtained from cultures grown in BHI broth to stationary phase ($OD_{600\ nm} \approx 1$) by a method based on a freeze-thaw procedure (Fagan and Fairweather 2011). Briefly, cultures of *C. difficile* were harvested by centrifugation at 5,000 x g for 10 min at 4°C and the pellets frozen at -20°C. Bacteria were thawed, suspended in PBS to an $OD_{600\ nm} = 20$ and incubated at 37°C for 10 min. Three such freeze-thaw cycles were carried out in order to obtain consistent and reproducible lysis.

The extraction of S-layer was performed following a previously described method (Fagan and Fairweather 2010).

To obtain S-layer-associated proteins of C. *difficile* strains, bacteria were grown in a 50 ml BHI broth for 16-18 h, harvested by centrifugation at 3500 x g for 10 minutes and washed

in 5 ml of PBS. Cells were re-suspended in 0.5 ml (1\100 starting culture) of 0.2 M Glycine, pH 2.2 plus protease inhibitors (Complete Mini EDTA-free, Roche) and digestion was allowed to proceed for 20 min at room temperature with gentle agitation. Then the bacterial suspension was transferred to microfuge tubes and harvested by centrifugation at 3500 x g for 10 minutes at 4°C. The supernatant, which contains the surface proteins, was recovered and the pH was modified by the addition of 2 M Tris base to achieve a pH in the range of 7–8.

For SDS-PAGE and Western blot analysis, 3.5 µl of each total cell extract and 5 µl of each S-layer extract were used.

Extraction of Flagella

C. difficile strains were inoculated in Falcon tubes containing 2 ml of BHI. Tubes were incubated overnight in an anaerobic hood. 250 µl of each culture were plated onto 4 plates. After incubation overnight, the colonies were collected from the plates with a loop and re-suspended in 500 µl of H_2O. All the samples were mixed by vortexing for 3 min in 500 µl of H_2O and centrifuged for 5 min at 13000 rpm at 4°C. The supernatant containing the flagella was transferred to a clean tube and used for SDS-PAGE analysis.

- **Expression of recombinant His-tagged proteins**

Selected clones were first grown in 25 ml LB medium with 100 µg/ml ampicillin overnight at 37°C with agitation at 180 rpm. The starter culture was then diluted to 500 ml of the same growth medium in a 2-liter flask and grown at 30°C with agitation at 180 rpm. When the culture reached an $O.D._{600\ nm}$ value ranging from 0.4 to 0.7, 1 mM IPTG (final concentration) was added for induction of expression. The culture was incubated at 25°C for 3.5 h with agitation at 180 rpm.

- **Protein expression using HTFS (High Throughput Fermentation System)**

This system is based on in-house adapted 50 ml Falcon tubes, prepared by boring three round holes in the cap: one in the centre for air intake and two lateral exit holes. The air intake hosts a 2 ml Falcon pipette connected to a fluximeter and the air exit is filtered by two shortened ART1000 filtered tips (Figure 17). Selected clones were then grown in 4 ml Glucose-HTCM at 37°C with agitation until bacteria reached the exponential phase. Thirty five microliters of each culture were inoculated into a solution containing 35 ml of Glycerol-HTCM and 50 µl PPC (poly propylene carbonate anti-foam solution, 1/10 diluted) in a

Falcon tube. The Falcon tubes were placed on a heater set to 26.1°C (to keep the temperature from dropping under 26°C) and air flux (which also provides agitation) was set to 3 VVM (volume gas / [volume liquid / minute]). The bacteria were grown for 36 h, then the Falcon tube caps are replaced with new ones and tubes were centrifuged at 4000 x g at 4°C for 40 min. Supernatants were discarded and pellets were stored at -20°C.

Figure 17. HTFS machinery has been developed in house at the Novartis Vaccines Fermentation Lab, by Erwin Swennen and colleagues. The system is composed of 2 block heaters holding a maximum number of 24 standard 50 ml Falcon tubes with 3 holes manually drilled in the cap. Filtered ART1000 tips are used as air-exits and the central hole is attached to a fluximeter set to 3 VVM

2.6 Protein purification.

- **Poly-Prep Column His-Tagged protein purification**

Pellets were thawed at room temperature and resuspended in 10 ml B-PER buffer (Bacterial- Protein Extraction Reagent, Pierce) containing 20 µl of 50 mM $MgCl_2$, 100µl DNAse I (100 K units Sigma D-4263) in PBS, 1 mg/ml lysozyme (Sigma L-7651) in PBS. Lysis solution was transferred to 50 ml centrifuge tubes, kept at room temperature for 40 min, vortexed a few times and then centrifuged at 40000 g for 25 min. Poly-Prep columns were prepared and equilibrated with 1 ml Ni-Activated Chelating Sepharose Fast Flow in 50 mM phosphate buffer and 300 mM NaCl at pH 8 before the supernatant was loaded. Column flow-through was discarded. Ten milliliters of 20 mM imidazole, 50 mM phosphate, 300 mM NaCl buffer at pH 8 were added to remove impurities.

Proteins bound to the column were eluted with 4.5 ml of 250 mM imidazole, 50 mM phosphate, 300 mM NaCl buffer at pH 8 and collected in three 1.5 ml fractions. 15µl of 200 mM DTT (Dithiothreitol) (2mM final concentration) were added to each fraction. Protein concentration of each fraction was estimated by the Bradford assay and 10 µg of each protein sample were loaded onto an SDS-PAGE gel. Proteins were then stored at 4°C. Poly-Prep Column GST-Tagged protein purification pellets were thawed at room temperature and

resuspended in 10 ml B-PER buffer (Bacterial-Protein Extraction Reagent, Pierce), 20 µl of 50 mM $MgCl_2$, 100 µl DNAse I (100 K units Sigma D-4263) in PBS, 1 mg/ml lysozyme (Sigma L-7651). Lysis solution was transferred to 50 ml centrifuge tubes, kept at room temperature for 40 min, vortexed a few times and then centrifuged at 40000 x g for 25 min. Poly-Prep columns were prepared, equilibrated with 0.5 ml Glutathione-Sepharose 4B, washed with 2 ml H_2O and 10 ml PBS, pH 7.4 before loading the supernatant. Column flow-through was discarded. 10 ml of PBS, pH 7.4 was added to remove impurities. Proteins bound to the column were eluted with 50 mM TRIS and 10 mM reduced glutathione at pH 8 and collected in three 1.5 ml fractions. 2 mM DTT was added to each fraction. Protein concentration of each fraction was estimated by the Bradford assay and 10 µg of each protein sample were loaded onto an SDS-PAGE gel.

2.7 SDS-PAGE and immunoblotting.

Extracts were separated by SDS-PAGE in a 12% polyacrylamide gel, followed by Western blotting and immunodetection with specific antibodies. Antisera against Cwp16, Cwp18, Cwp25, Cwp26 and Cwp27 were raised in mice immunised with purified recombinant CWPs obtained by overexpressing the corresponding ORFs of *C. difficile* strain 630 using the $pET15b^+$ vector (Novagen) and the *E. coli* strain BL21(DE3) (Invitrogen) expression system. Primary antibodies, used at a 1:2,000 dilution in blocking buffer, were detected using horseradish peroxidase-conjugated goat anti-mouse immunoglobulin G (1:20,000; Invitrogen) and the SuperSignal West Pico chemiluminescent substrate (Thermo Scientific Pierce). A marker for direct visualization of standard bands (MagicMark XP Western Protein Standard, Invitrogen) was used routinely for protein molecular mass estimation directly on Western blots.

2.8 Protein microarray

- **Design**

A total of 89 proteins were selected among those predicted, or known, to be secreted or surface exposed in *C. difficile*. Designing the chip layout is a key step in microarray-based research to overcome some limitations of the technique. For this experiment, we have designed a layout with four replicates of each protein (Figure 18). The same layout was then printed four times within the same slide, in order to obtain a microarray with a total of 16 replicates of each protein. The core of the Arrayjet technology, the JetSpyder, enables the inkjet print head to be simultaneously primed with multiple samples for microarray printing.

The design also included various types of controls, mostly serial dilutions of control proteins or immunoglobulins (at a concentration ranging from 0.008 to 0.5 mg/ml). A solution of 40% glycerol in PBS (PBS-Gly) buffer was spotted in at least twice the number of the protein spots and used to detect non-specific signals caused by cross-contamination during spotting (Figure 18).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	BSACy3/ Cy5 0,5 mg/ml	BSACy3/ Cy5 0,25 mg/ml	BSACy3/C y5 0,125 mg/ml	BSACy3/ Cy5 0,064 mg/ml	BSACy3/C y5 0,032 mg/ml	BSACy3/C y5 0,016 mg/ml	BSACy3/C y5 0,008 mg/ml	BSACy3/C y5 0,004 mg/ml	1	2	3	4	BSACy3/C y5 0,5 mg/ml	BSACy3/C y5 0,25 mg/ml	BSACy3/C y5 0,125 mg/ml	BSACy3/C y5 0,064 mg/ml	BSACy3/C y5 0,032 mg/ml	BSACy3/C y5 0,016 mg/ml	BSACy3/C y5 0,008 mg/ml	BSACy3/C y5 0,004 mg/ml	PBS_Gly	PBS_Gly	BSACy3/ Cy5 0,5 mg/ml
2	BSA_Biotin 0,5 mg/ml	BSA_Biotin 0,25 mg/ml	BSA_Biotin 0,125 mg/ml	BSA_Biotin 0,064 mg/ml	BSA_Biotin 0,032 mg/ml	BSA_Biotin 0,016 mg/ml	BSA_Biotin 0,008 mg/ml	BSA_Biotin 0,004 mg/ml	5	6	7	8	BSA_Biotin 0,5 mg/ml	BSA_Biotin 0,25 mg/ml	BSA_Biotin 0,125 mg/ml	BSA_Biotin 0,064 mg/ml	BSA_Biotin 0,032 mg/ml	BSA_Biotin 0,016 mg/ml	BSA_Biotin 0,008 mg/ml	BSA_Biotin 0,004 mg/ml	PBS_Gly	PBS_Gly	BSACy3/ Cy5 0,5 mg/ml
3	HuIgG 0,5 mg/ml	HuIgG 0,25 mg/ml	HuIgG 0,125 mg/ml	HuIgG 0,064 mg/ml	HuIgG 0,032 mg/ml	HuIgG 0,016 mg/ml	HuIgG 0,008 mg/ml	HuIgG 0,004 mg/ml	9	10	11	12	HuIgG 0,5 mg/ml	HuIgG 0,25 mg/ml	HuIgG 0,125 mg/ml	HuIgG 0,064 mg/ml	HuIgG 0,032 mg/ml	HuIgG 0,016 mg/ml	HuIgG 0,008 mg/ml	HuIgG 0,004 mg/ml	PBS_Gly	PBS_Gly	PBS_Gly
4	HuIgM 0,5 mg/ml	HuIgM 0,25 mg/ml	HuIgM 0,125 mg/ml	HuIgM 0,064 mg/ml	HuIgM 0,032 mg/ml	HuIgM 0,016 mg/ml	HuIgM 0,008 mg/ml	HuIgM 0,004 mg/ml	13	14	15	16	HuIgM 0,5 mg/ml	HuIgM 0,25 mg/ml	HuIgM 0,125 mg/ml	HuIgM 0,064 mg/ml	HuIgM 0,032 mg/ml	HuIgM 0,016 mg/ml	HuIgM 0,008 mg/ml	HuIgM 0,004 mg/ml	PBS_Gly	PBS_Gly	PBS_Gly
5	MoIgG 0,5 mg/ml	MoIgG 0,25 mg/ml	MoIgG 0,125 mg/ml	MoIgG 0,064 mg/ml	MoIgG 0,032 mg/ml	MoIgG 0,016 mg/ml	MoIgG 0,008 mg/ml	MoIgG 0,004 mg/ml	17	18	19	20	MoIgG 0,5 mg/ml	MoIgG 0,25 mg/ml	MoIgG 0,125 mg/ml	MoIgG 0,064 mg/ml	MoIgG 0,032 mg/ml	MoIgG 0,016 mg/ml	MoIgG 0,008 mg/ml	MoIgG 0,004 mg/ml	PBS_Gly	PBS_Gly	PBS_Gly
6	HaIgG 0,5 mg/ml	HaIgG 0,25 mg/ml	HaIgG 0,125 mg/ml	HaIgG 0,064 mg/ml	HaIgG 0,032 mg/ml	HaIgG 0,016 mg/ml	HaIgG 0,008 mg/ml	HaIgG 0,004 mg/ml	21	22	23	24	HaIgG 0,5 mg/ml	HaIgG 0,25 mg/ml	HaIgG 0,125 mg/ml	HaIgG 0,064 mg/ml	HaIgG 0,032 mg/ml	HaIgG 0,016 mg/ml	HaIgG 0,008 mg/ml	HaIgG 0,004 mg/ml	PBS_Gly	PBS_Gly	PBS_Gly
7	HuIgA 0,5 mg/ml	HuIgA 0,25 mg/ml	HuIgA 0,125 mg/ml	HuIgA 0,064 mg/ml	HuIgA 0,032 mg/ml	HuIgA 0,016 mg/ml	HuIgA 0,008 mg/ml	HuIgA 0,004 mg/ml	25	26	27	28	HuIgA 0,5 mg/ml	HuIgA 0,25 mg/ml	HuIgA 0,125 mg/ml	HuIgA 0,064 mg/ml	HuIgA 0,032 mg/ml	HuIgA 0,016 mg/ml	HuIgA 0,008 mg/ml	HuIgA 0,004 mg/ml	PBS_Gly	PBS_Gly	PBS_Gly
8	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
9	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76
10	77	78	79	80	NN-His	NN-His	NN-His	NN-His	NN-His	NN-His	NN-His	NN-His	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly
11	81	82	83	84	85	86	87	88	89	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	89	87	86	85	84	83	82	81
12	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	80	79	78	77	76	75	74	73
13	72	71	70	69	68	67	66	65	64	63	62	61	60	59	58	57	56	55	54	53	52	51	50	49
14	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29	28	27	26	25
15	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
16	BSA_Biotin 0,5 mg/ml	BSA_Biotin 0,25 mg/ml	BSA_Biotin 0,125 mg/ml	BSA_Biotin 0,064 mg/ml	BSA_Biotin 0,032 mg/ml	BSA_Biotin 0,016 mg/ml	BSA_Biotin 0,008 mg/ml	BSA_Biotin 0,004 mg/ml	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	BSA_Biotin 0,5 mg/ml	BSA_Biotin 0,25 mg/ml	BSA_Biotin 0,125 mg/ml	BSA_Biotin 0,064 mg/ml	BSA_Biotin 0,032 mg/ml	BSA_Biotin 0,016 mg/ml	BSA_Biotin 0,008 mg/ml	BSA_Biotin 0,004 mg/ml	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly
17	HuIgG 0,5 mg/ml	HuIgG 0,25 mg/ml	HuIgG 0,125 mg/ml	HuIgG 0,064 mg/ml	HuIgG 0,032 mg/ml	HuIgG 0,016 mg/ml	HuIgG 0,008 mg/ml	HuIgG 0,004 mg/ml	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	HuIgG 0,5 mg/ml	HuIgG 0,25 mg/ml	HuIgG 0,125 mg/ml	HuIgG 0,064 mg/ml	HuIgG 0,032 mg/ml	HuIgG 0,016 mg/ml	HuIgG 0,008 mg/ml	HuIgG 0,004 mg/ml	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly
18	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
19	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
20	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
21	73	74	75	76	77	78	79	80	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly
22	81	82	83	84	85	86	87	88	89	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	89	88	87	86	85	84	83	82	81
23	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	GST-NN	GST-NN	GST-NN	GST-NN	GST-NN	GST-NN	GST-NN	GST-NN	80	79	78	77
24	76	75	74	73	72	71	70	69	68	67	66	65	64	63	62	61	60	59	58	57	56	55	54	53
25	52	51	50	49	48	47	46	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31	30	29
26	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	HuIgA 0,004 mg/ml	HuIgA 0,008 mg/ml	HuIgA 0,016 mg/ml	HuIgA 0,032 mg/ml	HuIgA 0,064 mg/ml	HuIgA 0,125 mg/ml	HuIgA 0,25 mg/ml	HuIgA 0,5 mg/ml	28	27	26	25	HuIgA 0,004 mg/ml	HuIgA 0,008 mg/ml	HuIgA 0,016 mg/ml	HuIgA 0,032 mg/ml	HuIgA 0,064 mg/ml	HuIgA 0,125 mg/ml	HuIgA 0,25 mg/ml	HuIgA 0,5 mg/ml
27	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	HaIgG 0,004 mg/ml	HaIgG 0,008 mg/ml	HaIgG 0,016 mg/ml	HaIgG 0,032 mg/ml	HaIgG 0,064 mg/ml	HaIgG 0,125 mg/ml	HaIgG 0,25 mg/ml	HaIgG 0,5 mg/ml	24	23	22	21	HaIgG 0,004 mg/ml	HaIgG 0,008 mg/ml	HaIgG 0,016 mg/ml	HaIgG 0,032 mg/ml	HaIgG 0,064 mg/ml	HaIgG 0,125 mg/ml	HaIgG 0,25 mg/ml	HaIgG 0,5 mg/ml
28	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	MoIgG 0,004 mg/ml	MoIgG 0,008 mg/ml	MoIgG 0,016 mg/ml	MoIgG 0,032 mg/ml	MoIgG 0,064 mg/ml	MoIgG 0,125 mg/ml	MoIgG 0,25 mg/ml	MoIgG 0,5 mg/ml	20	19	18	17	MoIgG 0,004 mg/ml	MoIgG 0,008 mg/ml	MoIgG 0,016 mg/ml	MoIgG 0,032 mg/ml	MoIgG 0,064 mg/ml	MoIgG 0,125 mg/ml	MoIgG 0,25 mg/ml	MoIgG 0,5 mg/ml
29	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	HuIgM 0,004 mg/ml	HuIgM 0,008 mg/ml	HuIgM 0,016 mg/ml	HuIgM 0,032 mg/ml	HuIgM 0,064 mg/ml	HuIgM 0,125 mg/ml	HuIgM 0,25 mg/ml	HuIgM 0,5 mg/ml	16	15	14	13	HuIgM 0,004 mg/ml	HuIgM 0,008 mg/ml	HuIgM 0,016 mg/ml	HuIgM 0,032 mg/ml	HuIgM 0,064 mg/ml	HuIgM 0,125 mg/ml	HuIgM 0,25 mg/ml	HuIgM 0,5 mg/ml
30	PBS_Gly	PBS_Gly	PBS_Gly	PBS_Gly	HuIgG 0,004 mg/ml	HuIgG 0,008 mg/ml	HuIgG 0,016 mg/ml	HuIgG 0,032 mg/ml	HuIgG 0,064 mg/ml	HuIgG 0,125 mg/ml	HuIgG 0,25 mg/ml	HuIgG 0,5 mg/ml	12	11	10	9	HuIgG 0,004 mg/ml	HuIgG 0,008 mg/ml	HuIgG 0,016 mg/ml	HuIgG 0,032 mg/ml	HuIgG 0,064 mg/ml	HuIgG 0,125 mg/ml	HuIgG 0,25 mg/ml	HuIgG 0,5 mg/ml
31	BSACy3/ Cy5 0,5 mg/ml	BSACy3/ Cy5 0,25 mg/ml	BSACy3/C y5 0,125 mg/ml	BSACy3/ Cy5 0,064 mg/ml	BSA_Biotin 0,004 mg/ml	BSA_Biotin 0,008 mg/ml	BSA_Biotin 0,016 mg/ml	BSA_Biotin 0,032 mg/ml	BSA_Biotin 0,064 mg/ml	BSA_Biotin 0,125 mg/ml	BSA_Biotin 0,25 mg/ml	BSA_Biotin 0,5 mg/ml	8	7	6	5	BSA_Biotin 0,004 mg/ml	BSA_Biotin 0,008 mg/ml	BSA_Biotin 0,016 mg/ml	BSA_Biotin 0,032 mg/ml	BSA_Biotin 0,064 mg/ml	BSA_Biotin 0,125 mg/ml	BSA_Biotin 0,25 mg/ml	BSA_Biotin 0,5 mg/ml
32	BSACy3/ Cy5 0,5 mg/ml	BSACy3/ Cy5 0,5 mg/ml	PBS_Gly	PBS_Gly	BSACy3/C y5 0,004 mg/ml	BSACy3/C y5 0,008 mg/ml	BSACy3/C y5 0,016 mg/ml	BSACy3/C y5 0,032 mg/ml	BSACy3/C y5 0,064 mg/ml	BSACy3/ Cy5 0,125 mg/ml	BSACy3/ Cy5 0,25 mg/ml	BSACy3/ Cy5 0,5 mg/ml	4	3	2	1	BSACy3/C y5 0,004 mg/ml	BSACy3/C y5 0,008 mg/ml	BSACy3/C y5 0,016 mg/ml	BSACy3/C y5 0,032 mg/ml	BSACy3/C y5 0,064 mg/ml	BSACy3/Cy 5 0,125 mg/ml	BSACy3/C y5 0,25 mg/ml	BSACy3/ Cy5 0,5 mg/ml

Figure 18 Microarray layout. Spots of *C. difficile* surface proteins are represented by white boxes, PBS_Gly buffer spots are highlighted in blue.

- **Preparation and Spotting**

Purified *C. difficile* proteins were dialyzed in PBS with the Slide-A-Lyzer Dialysis Cassette system (Thermo Scientific) and diluted to obtain a concentration of 0.5 mg/ml.

Each protein solution was then transferred to two 384-well polypropylene microplates (15 µl/well). The plates contained serial dilutions, ranging from 0.5 mg/ml to 0.016 mg/ml, of seven controls: Human IgG, Human IgM, Human IgA, Mouse IgG, Hamster IgG, Cy3 and Cy5-labelled BSA (Amersham Biosciences) and biotin-labelled BSA. All samples were spotted onto nitrocellulose-coated slides by using the Arrayjet Marathon Microarrayer.

The Arrayjet Marathon Microarrayer transfers samples of biological liquids from microplates onto microarray slides. The samples are transferred by means of an inkjet print head and a patented JetSpyderTM. The JetSpyderTM allows 12 or 32 samples to be drawn from wells into the print head as follows: the print head is purged with clean buffer liquid (50% water, 50% glycerol and 0.05% Triton X100); the print head picks up the JetSpider, which is then purged with buffer via the print head; the Jet Spider, with 12 capillaries projecting downwards in a 4x3 array at 9 mm spacing, or 32 in a 8x4 array at 4.5 mm spacing, is lowered into a microtitre plate. Samples are drawn from the wells via the JetSpyder into the nozzles of the print head; the JetSpider is returned to its cleaning station; the print head prints the slides and the cycle is repeated until all the samples in the plates are printed. The Jet Spider draws 0.7 µl of sample and each nozzle prints 0.1 nl/spot (0.05 ng of protein/spot) for a final spot diameter of 90–100 µm.

The wells must be filled to a level adequate for the planned printing, and must not contain dirt or air bubbles. Centrifuging the plates at 3000 rpm for 5 minutes should remove bubbles, but the plates should be inspected and re-centrifuged if necessary. The software allows the user to specify a print run; initialise the instrument; load plates and slides; perform the print run; unload trays and slides; and shut down the instrument. The software will produce a GAL file describing the mapping of the contents of the wells onto spots on the slides. The GAL file is needed to map the positions of the samples in the plates to the positions of the corresponding spots on the slides.

- **Hybridization and Staining**

The slides were washed with PBS at 20°C for 5 min and then with PBS-T (0.05% Tween 20 in PBS) at 20°C for 1 min followed by 1 h incubation in the dark with shaking in Protein Blocking Buffer (Whatman code: 10485356). Slides were incubated with 100 µl of hamster sera diluted 1:300 in Protein Blocking Buffer for 1 h at 25°C in the dark with shaking and then washed 3 times (at 25°C, 5 min each time) with PBS-T. Anti-hamster IgG/Alexa Fluor647 secondary antibody (Invitrogen, cod.A21451) at a 1:800 dilution in

Protein Array Blocking buffer, was added and incubation prolonged for 1 h at 25°C in the dark with shaking. Only in the last slide (00282619) an Anti-hamster IgG/Cy3 (abcam, code: ab6969), diluted 1:800 in Protein Array Blocking Buffer was added. Slides were washed twice with PBS-T (25°C, 5 min each time), once with PBS (25°C, 10 min), and once with sterile milliQ H_2O (25°C, 30 sec). Slides were removed from the incubation chamber, washed once with sterile milliQ H_2O and dried with nitrogen.

An example of the scheme used for a hybridization experiment of the 8 hamster sera analysed in this work is given below

	No serum	1	2	3	4
Serum dilution	-------	1:300	1:300	1:300	1:300
Anti-HalgG Alexa Fluor647	1:800 Anti-HalgG/ Alexa Fluor647	1:800 Anti-HalgG/ Alexa Fluor647	1:800 Anti-HalgG/ Alexa Fluor647	1:800 Anti-HalgG/ Alexa Fluor647	1:800 Anti-HalgG/ Alexa Fluor647
slide	00282460	00282461	00282462	00282611	00282612

	5	6	7	8	4
Serum dilution	1:300	1:300	1:300	1:300	1:300
Anti-HalgG Alexa Fluor647	1:800 Anti-HalgG/ Alexa Fluor647	1:800 Anti-HalgG/ Alexa Fluor647	1:800 Anti-HalgG/ Alexa Fluor647	1:800 Anti-HalgG/ Alexa Fluor647	1:800 Anti-HalgG/ Cy3
slide	00282614	00282615	00282616	00282617	00282619

- **Data collection**

The fluorescence signals were detected using a high resolution laser scanner PowerScanner™ (Tecan) to detect the specific fluorophore used, with 0.1% laser intensity and 300% of gain for all the slides, except for the last slide (00282619) in which a 0.1% laser intensity and 90% of gain were used.

The signal was then quantified with the program ImaGene 9.0 (Biodiscovery Inc, CA, USA). The data collected were analysed using the in-house developed program "Protein Chip". The mean fluorescence intensity (MFI) of each spot was calculated as an average fluorescence intensity of the 16 replicates, minus the background fluorescence intensity. To determine specific binding of an antibody-probe to a recombinant protein, a mean fluorescence intensity value of 2000 was established as the threshold for scoring positive results. Average intensities of fluorescence below this value were scored as negative.

3. RESULTS

3.1 Sequence conservation of *cwp* genes

Sequence alignment of ORFs predicted to code for CWPs in published *C. difficile* genomes (Table 2) was used to identify conserved flanking regions suitable for designing primers for the amplification of the corresponding *cwp* ORFs.

Table 2. List of published *C. difficile* genomes, available until 2011.

C. difficile strain	institute	ribotype
630	Sanger Institute	012
CD196	Wellcome Trust	027
R20291(SM)	Wellcome Trust	027
QCD-97b34	McGill University	-
QCD-76w55	McGill University	-
QCD-66c26	McGill University	-
QCD-63q42	McGill University	-
QCD-37x79	McGill University	001
QCD-32g58	Washington University	-
QCD-23m63	McGill University	078
CIP107932	McGill University	-
ATCC43255(VPI10463)	McGill University	-
BI-1	Wellcome Trust	027
M120	Wellcome Trust	078
M68	Wellcome Trust	017
CF5	Wellcome Trust	017
855	Wellcome Trust	027
gs	University of Arizona	-
NAP08	BCM	-
NAP07	BCM	-

Among the public genomes, we selected three fully sequenced genomes to use as reference in our analysis. The genomes chosen correspond to the strains 630, R20291 (SM) and M120. Strain 630 (ribotype 012) is multi-drug resistant, and was isolated from a patient with severe pseudomembranous colitis that had spread to dozens of other patients in the same ward in Zurich, Switzerland in 1982. Subsequently, strain 630 has become the most widely used strain in the laboratory. The strain R20291 (SM) was isolated in the UK in 2006. This strain (PCR ribotype 027) is characterised by high level fluoroquinolone resistance, hyper-production of the pathogenic toxins A and B and has been implicated in the increase of incidence, recurrence and mortality of *C. difficile* associated disease (CDAD) in hospitalised patients. The strain M120, a PCR ribotype 078 strain, was isolated from an Irish diarrheic patient in 2007. The PCR ribotype 078 strains are generally those most frequently present in domestic animals.

Twelve of the 29 *cwp* genes encoding the cell wall proteins are clustered in the same region of the genome, named after *slpA* (*cwp1*) the *slpA* locus (Calabi, Ward et al. 2001; Karjalainen, Waligora-Dupriet et al. 2001), whereas the remaining 17 paralogs are distributed throughout the genome (Figure 19). We conducted our analysis on all genes not included in the *slpA* locus, as this locus is already been characterized by other groups (Karjalainen, Saumier et al. 2002; Eidhin, Ryan et al. 2006; Kirby, Ahern et al. 2009; Chapeton Montes, Collignon et al. 2013).

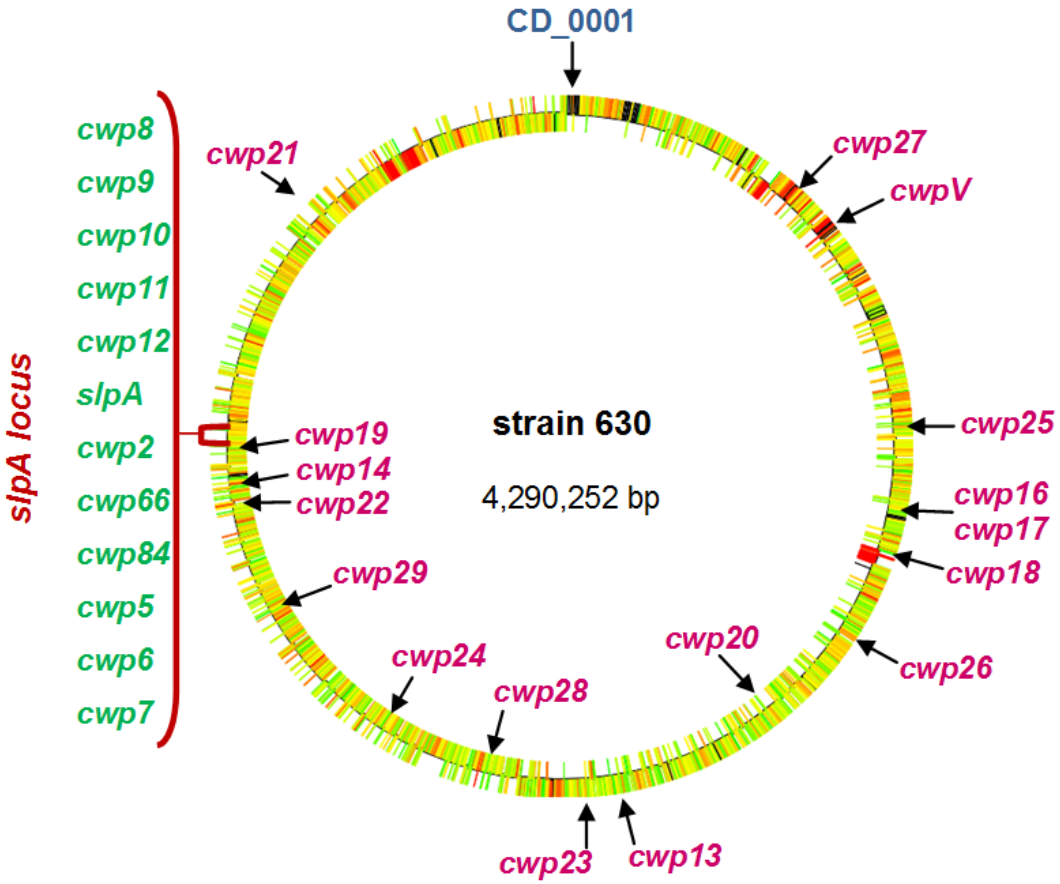


Figure 19.Distribution of *cwp* genes in the genome of *C.difficile*. Genes in pink represent the genes analysed in our study. Only the 17 *cwp* ORFs not included in the slpA locus were taken into consideration.

In several cases only one of the two flanking regions was conserved in all the published genomes. In these cases, different sets of primers were designed that would allow the amplification of the corresponding ORF in all the known variants of the locus. A list of the primers used is given in Table 3.

Table 3.Oligonucleotides used in this study

Oligonucleotide	Sequence (5'-3')
cwp13-for	AAGGGGGAGAAAGCGTGAAAA
cwp13-rev	AAACTTATACTACATAAATTTAAAGCTG
cwp13-for1	GTCCAACAAAAAGTTTTAATACAAATC
cwp13-rev1	GATATATTTAAGCTTTTTAATTGTTGTTC
cwp13-for2	GTATATTTGTAAATACTTGATATTTAGTC
cwp15-for1	CTTAATAGGTGGAACTTCTGTATTAA
cwp15-rev2	ATCATCTATAACTACATCATCCAC
cwp16-for	GGCAATAAATGATATAGTGAAGATAA
cwp16-rev	TAATTACATTCTTATCTTATAGCTTTAGT
cwp16-for1	GAAAAATGGTTCAGATAAGGTAG
cwp17-for	GGGTTAGATACACTTTGACATATT
cwp17-rev	CAAATGTACATACAGTCTTACTATTG
cwp17-for1	GAAACAGCAGTTAAAATAAGTAAAGA
cwp18-for	GCTATATCATTCTTTATAAGTCAAGC
cwp18-rev	CTATATTGAAATATAACTTAGTTTACTTG
cwp19-for	TACACTTGATTACAAAATGCATATAG
cwp19-rev	TACACTTGATTACAAAATGCATATAG
cwp19-for1	ACCCAGGTTTACCAGAAGTTA
cwp19-rev1	CAAGTAAAGTTATTTTAATTGGAGGA
cwp20-for	ATTATTAAAAAGGGAAAGAGGTAATG
cwp20-rev	GTCATAGAGCTATGGGAGCTA
cwp20-for1	GTTATATAAAAAGCTATACTTATAAAAGC
cwp20-for2	GAGTATCTGGTATGTCATTTACTAA
cwp20-rev1	GATGCAGTAAGCATTGGAGCT
cwp22-for	AAGATTCATGGAGATGGCAAAG
cwp22-rev	GCTTTTATTGATTGAATTTCACATCT
cwp24-for	AGCGTAGATGTAATACTATGTTA
cwp24-rev	GAAATTGGAAAAATCAAAACCATTCC
cwp25-for	CCTTTTATGGAGAAAGAAGGTATT
cwp25-rev	CTTATCTCCATTTGATAGCCTC
cwp26-for	ATAATTTAGGTGAAAATATTTAAGAGAAG
cwp26-for1	GAAAAAATAGAGTGGACACCTCATT
cwp26-rev	ATGAACTCCACTCTATAGATTGTA
cwp27-for	GTGACAAAATTTTGAAATATAAACTAATTTC
cwp27-rev	GAAACATAAAATTTTAATACTTATCGTC
cwp28-for	GCAATTTTACATAAGTTTGTTAAGTAC
cwp28-rev	CCTCCTAGTTAATTTCTTATATTACAC
cwp29-for	GGAGGAAAAGAAAATTTAATGTAGAAT
cwp29-rev	GAGACATTTACATACCACATAGT

All the sets of primers whose sequence was highly conserved in all published genomes or, alternatively, could discriminate for the presence/absence of a specific *cwp* gene were combined as two sets in a multiplex PCR reaction and tested on control genomic DNAs extracted from strains 630, R20291 and M120. With this approach, amplification of a conserved PCR fragment becomes a positive control for the negative result obtained when using primers specific for genomic regions that are not present in all published genomes (Figure 20).

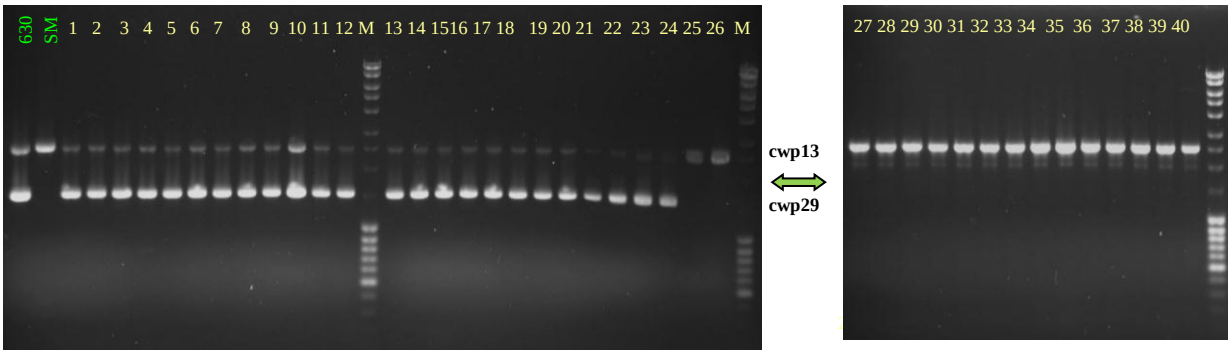


Figure 20. . Results from duplex PCR amplification of loci *cwp13* and *cwp29* performed on the clinical isolates of Table1

PCR fragments of the expected length were obtained with all sets of primers, except for those designed for the amplification of *cwp14*, *cwp21* and *cwp23*. Hence, these genes were excluded from our analysis that was focused on the remaining 14 *cwp* genes listed in Figure 22. Amplification of the corresponding 14 *cwp* ORFs was carried out on genomic DNA extracted from 40 strains isolated from patients at 13 Italian hospitals and representative of the 6 PCR ribotypes prevalent in Italy (Table 1). The nucleotide sequence was then determined for all of the PCR fragments obtained (Figure 21).

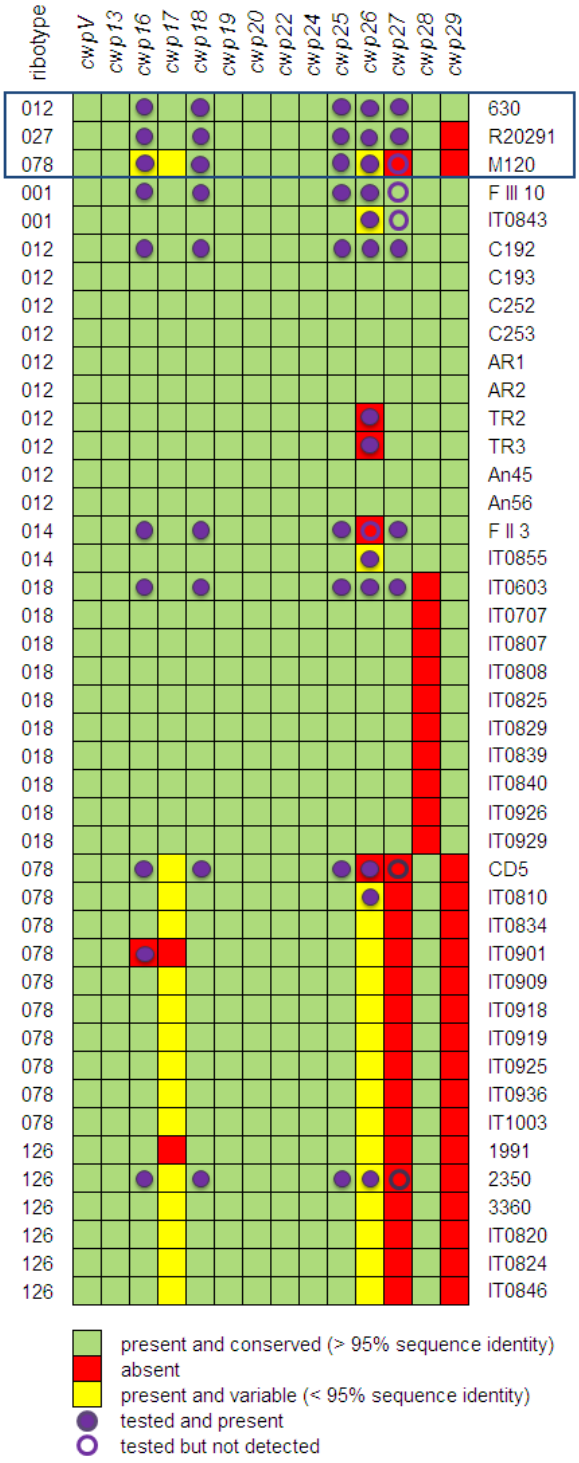


Figure 21 Distribution of the 14 *cwp* genes in 40 Italian clinical isolates and their conservation with respect to strain 630. Reference strains are highlighted in the blue box. Strains also tested for expression of a specific Cwp by Western blot analysis are indicated by a full circle (●) when found positive or by an open circle (O) when a specific band could not be detected.

In total, 511 ORFs were sequenced and analysed for sequence conservation using the multiple alignment program Clustal W (Chenna, Sugawara et al. 2003). The ORFs were found to be conserved in all 40 isolates for eight of the *cwp* genes analysed, while the

remaining six *cwp* ORFs were absent or, when present, not conserved in at least one of the clinical isolates. In particular, *cwp27* and *cwp29* were absent in all the PCR ribotypes 078/126 isolates as already reported for M120, the reference strain for PCR ribotype 078 (He, Sebaihia et al. 2010). Similarly, the ten PCR ribotype 018 strains of our collection lack *cwp28*, while in the same isolates the other 13 *cwp* genes were found to be present and conserved (Figure 21). A PCR fragment of the expected length for *cwp17* was obtained in all isolates except for two, one PCR ribotype 078, which lacks also *cwp16*, and one PCR ribotype 126 strain.

Moreover, the sequence of *cwp17* in all the remaining isolates of PCR ribotype 078 and 126 showed a lower level of conservation with respect to strain 630 than the isolates of the other PCR ribotypes. Finally, the *cwp26* gene was found to be alternatively absent/present but variable or conserved in different isolates of the same PCR ribotype with the exception of the ten PCR ribotype 018 isolates, which all shared identical *cwp26* sequences. The number of single nucleotide polymorphisms (SNPs) found in each *cwp* gene among the various PCR ribotypes with respect to the ortholog sequence in strain 630 and the corresponding amino acid substitutions are reported in Table 4.

Table 4.Gene variability in 14 *cwp* genes of *Clostridium difficile*. The sequence determined in a representative clinical isolate for each PCR ribotype was compared to the equivalent sequence in strain 630 by measuring the number of SNPs and amino acid substitutions observed.

		cwpV		cwp13		cwp16		cwp17		cwp18		cwp19		cwp20		cwp22		cwp24		cwp25		cwp26		cwp27		cwp28		cwp29	
PCR-ribotype	Reference strain	SNPs (bp)	Substitutions (a.a.)	SNPs (bp)	Substitutions (a.a.)	SNPs (bp)	Substitutions (a.a.)	SNPs (bp)	Substitutions (a.a.)	SNPs (bp)	Substitutions (a.a.)	SNPs (bp)	Substitutions (a.a.)	SNPs (bp)	Substitutions (a.a.)	SNPs (bp)	Substitutions (a.a.)	SNPs (bp)	Substitutions (a.a.)	SNPs (bp)	Substitutions (a.a.)	SNPs (bp)	Substitutions (a.a.)	SNPs (bp)	Substitutions (a.a.)	SNPs (bp)	Substitutions (a.a.)	SNPs (bp)	Substitutions (a.a.)
027(R20291)	630	7	1	25	7	28	6	18	9	1	-	8	4	32	12	7	4	2	-	5	-	9	5	7	3	4	1		
078(M120)	630	10	-	97	34	180	66	140	46	35	11	54	38	118	32	50	20	39	4	27	6	154	75			52	14		
001	630	4	-	9	2	22	3	13	6	10	5	3	1	36	11	5	1	2	-	2	-	6	3	6	2	2	-	12	6
012	630	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
014	630	5	-	6	2	18	5	13	7	3	2	3	2	32	9	5	1	3	-	5	-	9	5	4	2	26	10	12	6
018	630	5	-	4	1	21	4	12	6	2	1	3	2	35	11	5	1	-	-	6	-	12	6	5	2			12	6
078	630	10	-	97	34	86	18	140	46	35	11	54	38	118	32	50	20	39	4	27	6	154	75			52	14		
126	630	10	-	97	34	86	18	140	46	35	11	54	38	118	32	50	20	39	4	27	6	154	75			52	14		

The sequences of the 14 *cwp* ORFs in one isolate for each PCR ribotype, arbitrarily selected as representative of all strains that belong to the same PCR ribotype, were joined in a single string and compared with the corresponding cwp sequences of strain 630, also joined in a string.

The phylogenetic tree inferred from the sequences of the 14 *cwp* loci among the clinical isolates and constructed by the use of the neighbour-joining algorithm is shown in Figure 22. Two interesting observations can be drawn from this analysis. First, strains that belong to PCR ribotypes 078 and 126 always have identical *cwp* sequences in the clinical isolates of our collection. Conversely, the reference strain for PCR ribotype 078 M120 and strain QCD-23M63, another PCR ribotype 078 strain whose genome sequence has been characterised (Forgetta, Oughton et al. 2011), show some variability for these *cwp* genes (Figure 23). Second, the two PCR ribotypes 078 and 126 are clearly more closely related to the hypervirulent PCR ribotype 027 than to any of the other PCR ribotypes analysed in our study.

.

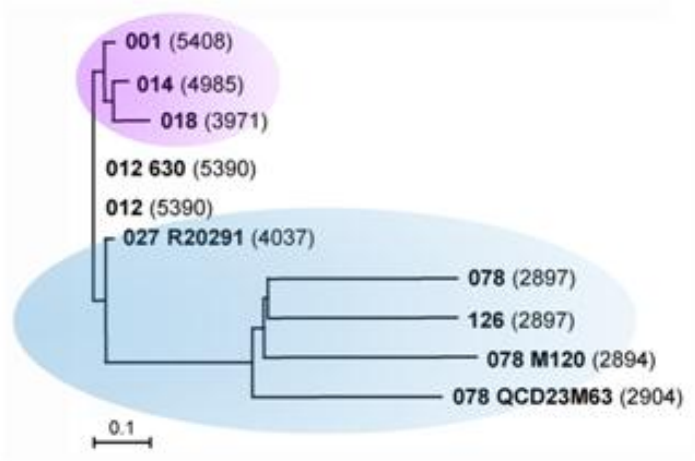


Figure 22. Phylogenetic tree, rooted on *C. difficile* strain 630, of *cwp* sequences among *C. difficile* clinical isolates. The tree was constructed from the sequence results inferred from 14 *cwp* gene sequences by using the neighbour-joining algorithm. Bootstrap confidence values for each node of the tree were calculated over 100 replicate trees (only bootstrap values >80% are indicated). The number of nucleotides representing all polymorphisms found in each PCR ribotype is given in parentheses.

3.2 Analysis of *in vitro* expression of Cwp proteins

Five of the CWPs under study were characterised further by Western blot analysis of total cell extracts or S-layer preparations of the *C.difficile* clinical isolates. Three conserved CWPs, Cwp16, Cwp18 and Cwp25, and two variable ones, Cwp26 and Cwp27, were selected for this analysis since antisera showing high specificity for these Cwps were available. For simplicity, total cell extracts were prepared from only one representative isolate for each PCR ribotype as well as from any strain showing variable alleles. A similar mode of expression was detected in total extracts of all the clinical isolates analysed for the highly conserved *cwp16, cwp18* and *cwp25* genes (Figure 23).

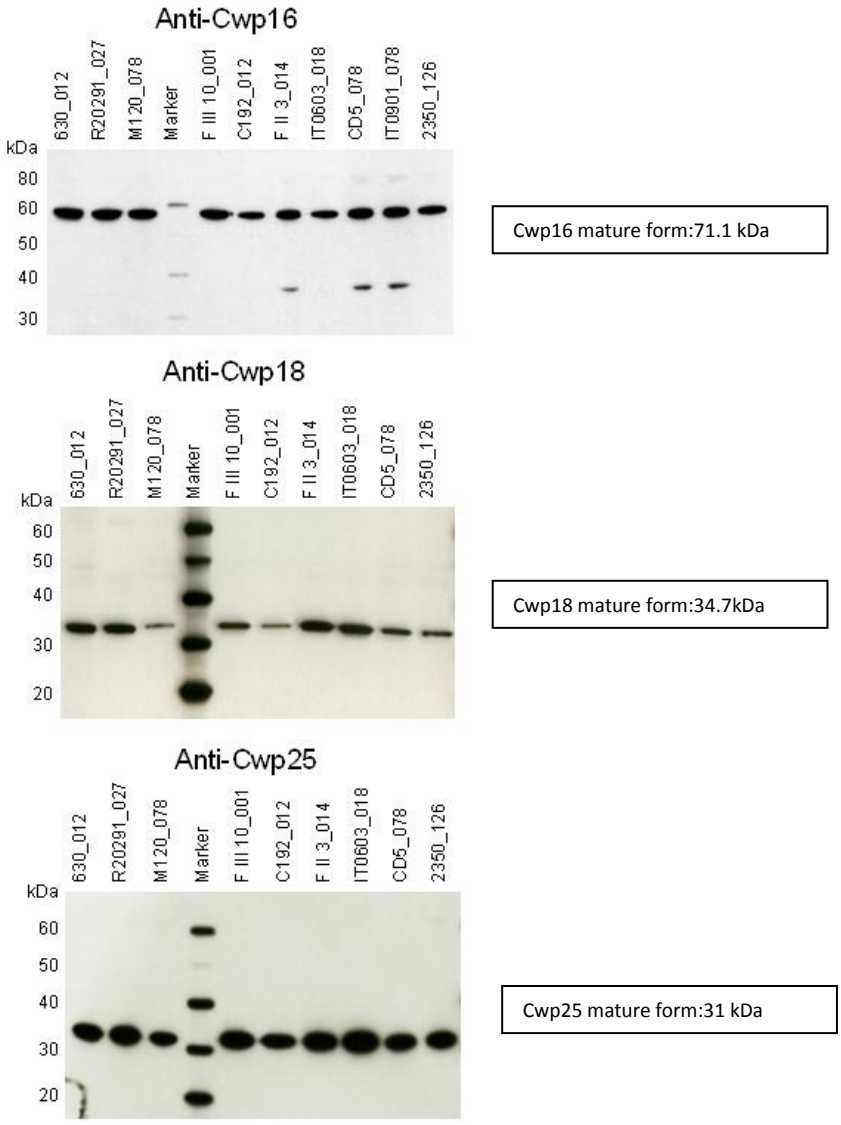


Figure 23. Analysis of expression of conserved CWPs. Western blot analysis of total cell extracts of *C. difficile* reference strains (630, R20291 and M120) and clinical isolates representing different PCR ribotypes using anti-Cwp16, anti-Cwp18 and anti-Cwp25 antibodies. A marker for direct visualization of standard bands (MagicMark XP Western Protein Standard, Invitrogen) was used for protein molecular mass assessment directly on Western blots.

However, it is noteworthy that strain IT0901, the only isolate from which we were not able to amplify the *cwp16* gene, shows a Cwp16-positive band of the same intensity and molecular mass as all the other strains. This suggests that the *cwp16* flanking regions that we used to design the primers are not conserved in strain IT0901. Moreover, since we were also not able to amplify the adjacent *cwp17* gene in this isolate, we propose that in strain IT0901 the entire region may contain some degree of sequence variability that does not compromise expression of Cwp16. For this reason, we believe Cwp16 could be included in the group of the highly conserved CWPs, thus bringing to 9 the number of conserved CWPs versus 5 variable ones. In the literature there is a paucity of information on Cwps. To date, only SlpA, CwpV, Cwp84 and Cwp13 have been characterised. The main aspect that they all have in common is that they are processed through a complex pathway before exposure on the bacterial surface. We believe that Cwp 16 is subjected to the same processing, as a band is visible at a lower molecular weight in all the samples.

Conversely, analysis of the data obtained on expression of Cwp26 revealed that a protein of the expected molecular mass (49 kDa) is present in total cell extracts of the 027 reference strain R20291 and the PCR ribotype 001 strains, but missing in the remaining isolates (Figure 24 A). A weaker band visible at approximately 31 kDa in all samples, representing a cross-reaction of the Cwp26-specific polyclonal antibody with Cwp25 (data not shown), was used as a sample loading control for the Cwp26-negative samples. To verify if the absence of Cwp26 in total cell extracts was due to the sample preparation procedure or to differences in expression/localisation, the Western blot analysis was repeated on S-layer preparations of the same strains. S-layer extracts showed the presence of a 49 kDa Cwp26-positive band in the PCR ribotype 078 and 126 strains as well as in the 027 and 001 isolates already found to be positive in total cell extracts (Figure 24 B).

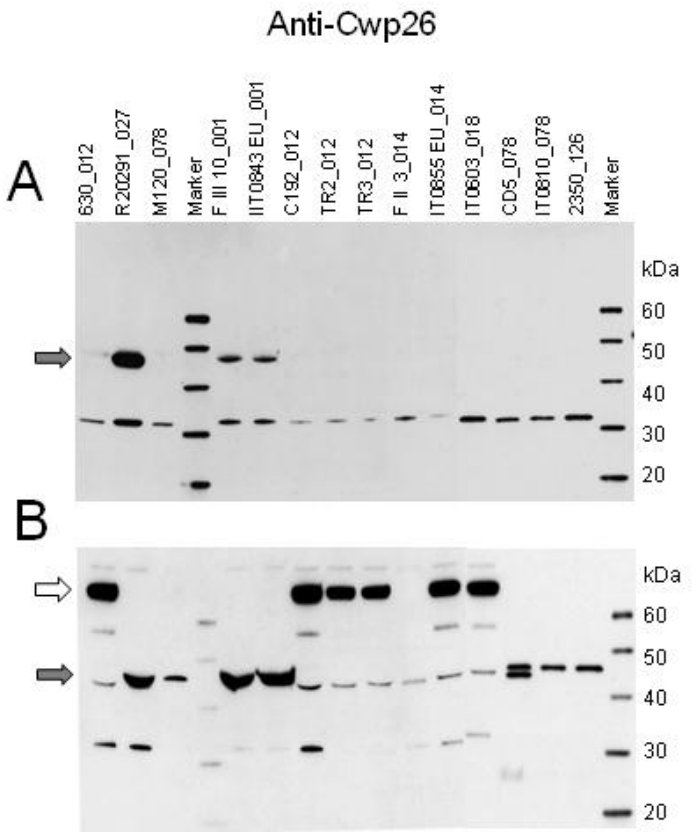


Figure 24. Expression of Cwp26. Western blot analysis of *C. difficile* reference strains (630, R20291 and M120) and clinical isolates representing different PCR ribotypes using anti-Cwp26 antibodies. Total cell extracts (**A**) and S-layer extracts (**B**) were separated by SDS-PAGE, followed by Western blotting with Cwp26-specific antibodies. A marker for direct visualization of standard bands (MagicMark XP Western Protein Standard, Invitrogen) was used for protein molecular mass assessment directly on Western blots. The grey arrow indicates the position of the expected molecular mass for the mature form of Cwp26, while the white arrow indicates the Cwp26-positive band at approximately 70 kDa.

Likewise, the strains that belong to the PCR ribotypes 012, 014 and 018 showed a strong positive signal only in S-layer preparations, though the strong band recognised in these strains has a significantly higher molecular mass (approx. 70 kDa) than that predicted from the *cwp26* gene sequence (Figure 24B). In addition, it should be noted that the same strong signal at 70 kDa is also visible in S-layer preparations of the two 012 isolates, TR2 and TR3, that were found *cwp26*-negative by PCR analysis (Figure 21). Of the other two *cwp26*-negative isolates reported in Figure 1, strain CD5 displayed two positive bands at 48 and 49 kDa, while strain F II 3 did not show any specific band recognised by the anti-Cwp26 antibodies (Figure 24 B). Although we cannot offer an explanation for the results obtained in isolates TR2, TR3 and CD5, it can be inferred that F II 3 is the only strain that clearly does not contain a *cwp26* ortholog. The results of the Western blot analysis of the expression of Cwp27 are shown in Figure 25.

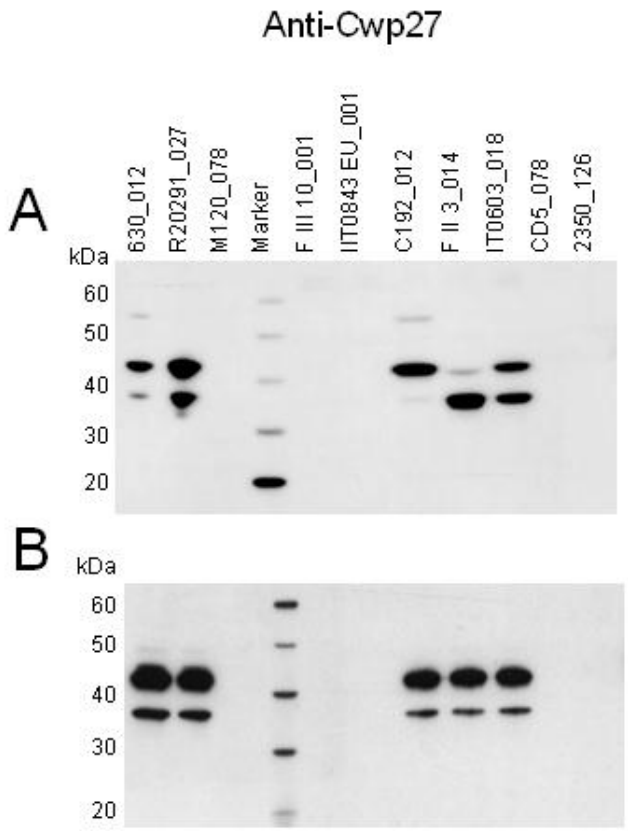


Figure 25. Expression of Cwp27. Western blot analysis of *C. difficile* reference strains (630, R20291 and M120) and clinical isolates representing different PCR ribotypes using anti-Cwp27 antibodies. Total cell extracts (**A**) and S-layer extracts (**B**) were separated by SDS-PAGE, followed by Western blotting with Cwp27-specific antibodies. A marker for direct visualization of standard bands (MagicMark XP Western Protein Standard, Invitrogen) was used for protein molecular mass assessment directly on Western blots.

Two bands of the expected molecular mass for the mature form (38 kDa) and precursor (41 kDa) of Cwp27, as inferred from the Cwp27 sequence analysis using the PSORT (Nakai and Horton 1999) and VECTOR NTI prediction programs, were seen in total extracts as well as in S-layer preparations of the two reference strains 630 and R20291 used as positive controls. Conversely, no band was visible in M120, a strain that does not have a *cwp27* ortholog.

In clinical isolates of PCR ribotype 012, 014 and 018 the same two bands were present with the same intensity ratio in S-layer extracts but with varying intensities in total cell extracts. However, both the PCR ribotype 001 isolates of our collection, which contain a conserved *cwp27* gene, did not show any positive signal; thus indicating that these strains do not express Cwp27 at detectable levels in our conditions. On the contrary, the lack of a positive signal in PCR ribotype 078 and 126 isolates confirms the absence of a *cwp27* ortholog in these strains.

3.3 Analysis of *in vivo* expression by Protein Microarray

Reverse Vaccinology highlights how antigens that confer a broad range of protective antibody responses are highly expressed surface-exposed proteins, usually secreted as toxins or virulence factors well conserved among the pathogen strains. The identification of antigens able to confer protective immunity can be achieved with several approaches. In this study, we chose to analyse the immune response obtained in a protection experiment, carried out in hamsters, using a protein chip approach. Briefly, hamsters were vaccinated with various domains of Toxin A (p5 /6) and Toxin B and then challenged with a lethal dose of *C. difficile* strains 630 or B1. We have used the sera from protected and from control animals to identify antigens capable of eliciting a specific immune response towards *C. difficile* toxins and surface proteins. Putative antigens or proteins predicted to be secreted or surface exposed by genomic and proteomic approaches were purified in a His-tagged form, and used to generate protein microarrays on nitrocellulose-coated slides. Approximately 80 proteins were selected by using a combined *in silico* and proteomic approach (Table 5).

Table 5. List of proteins spotted onto the chip.

Tube N°	Name	note
1	A 16 small His	ToxA+ToxB_1
2	B4 His	ToxA+ToxB_2
3	Chimera 9 His	ToxA+ToxB_3
4	Chimera 12 His	ToxA+ToxB_4
5	N-G-BB His	N-G-BB His
6	PTA2 His	ToxA_1
7	Tox A-B2 His	ToxA_2
8	Tox A-GT-WT	ToxA_3
9	Tox A-GT-M2	ToxA_4
10	Tox A-GT-M3	ToxA_5
11	Tox B-GT His	ToxB_1
12	Tox B1 His	ToxB_2
13	Tox B4 His	ToxB_3
14	ToxB4	ToxB_4
15	FliC_630	FliC_630 (native)
16	FliC_SM	FliC_SM (native)
17	FliC_001	FliC_001 (native)
18	FliC_014	FliC_014 (native)
19	FliC_018	FliC_018 (native)
20	FliC_078	FliC_078 (native)
21	FliC_126	FliC_126 (native)
22	HMW_630	HMW_630 (native)
23	HMW_SM	HMW_SM (native)

24	DIF 2 His	Leader present, peptidoglycan catabolic process, putative cell wall protein (no lpxtg)
25	Dif 14 B His (domain)	flagellar cap protein FliD family
26	DIF 15 His	No leader, PSORT:extracellular, flagellar protein
27	DIF 16 His	Lipoprotein leader
28	20f30-563 His	Leader present, lpxtg present, cna_B domains
29	20f560-975 His	Leader present, lpxtg present, cna_B domains
30	DIF SleC His	SleC His
31	DIF 40 His	Leader present, fimbrial protein
32	DIF 44 His	Leader present, cell wall-binding domain **cwp25**
33	DIF 51 His	Lipoprotein signal
34	DIF 52 His	Leader present, bruttina, lpxtg like signal present
35	Dif 53 His	**cwp18**
36	Dif 55 His	putative cell wall hydrolase
37	DIF 75A His	Leader present, cell wall binding domain present **cwp20**
38	Dif 89 His	Lipoprotein signal
39	Dif 104 His	Lipoprotein signal
40	Dif 106C	**cwp24**
41	DIF 109B His	No leader, PSORT:extracellular
42	DIF 114 His	Lipoprotein signal
43	Dif 130 His	Lipoprotein signal
44	Dif 139 His	**cwp14**
45	Dif 144 His	**cwp7**
46	Dif 145 His	**cwp6**
47	Dif 146 His	**cwp5**
48	Dif 149 His	**cwp11**
49	Dif 167 His	Leader present
50	Dif 170 His	Leader present, NLPC_P60 domain
51	DIF 171 A His	Leader present, lpxtg like motif, cna_B domains, 95% identity with DIF20
52	DIF 171 B His	Leader present, lpxtg like motif, cna_B domains, 95% identity with DIF20
53	Dif 173 His	Leader present
54	Dif 184 His	No leader, collagen binding protein
55	Dif 187 His	**cwp27**
56	Dif 189 A His	Leader present, cw binding domains **cwpV**
57	Dif 189 A (dominio)	Leader present, cw binding domains **cwpV**
58	Dif 189 B His	Leader present, cw binding domains **cwpV**
59	DIF 192 His	Leader present, cw binding domains **cwp16**
60	Dif 194 His	Signaling protein
61	DIF 196 His	Leader present, cw binding domains **cwp13**
62	Dif 201His	**cwp29**
63	Dif 204 His	**cwp66**
64	Dif 205A	**SlpA (from 630) recombinant**
65	DIF 207 His	Leader present, cw binding domains **cwp10**
66	DIF 208A His	Leader present, lpxtg like motif, cna_B domains
67	DIF 208B His	Leader present, lpxtg like motif, cna_B domains
68	DIF 210 His	Leader present, lpxtg like motif
69	DIF 211 His	Leader present, cw binding domains **cwp21**
70	DIF 212 His	Leader present, lpxtg like motif

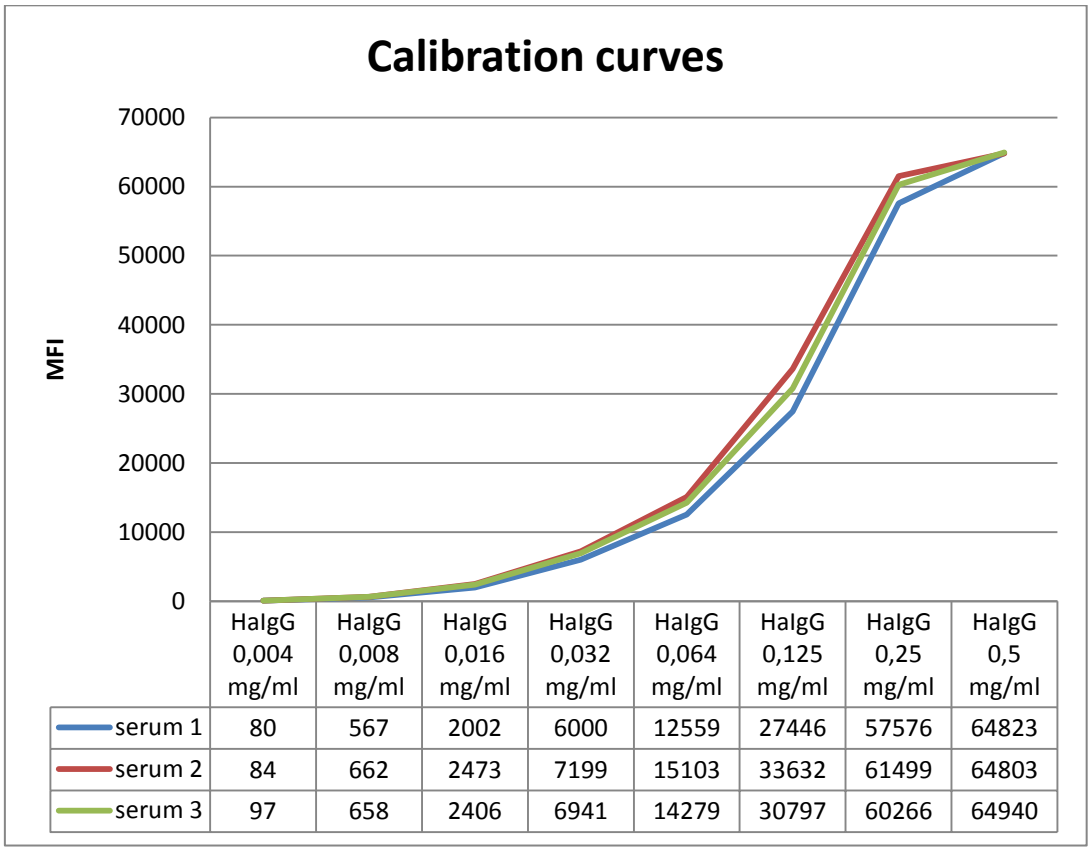
71	DIF 225 His	Leader present, lpxtg like motif
72	Dif 227 His	Leader present
73	DIF 231 His	Leader present, lpxtg like motif
74	DIF 232 His	Leader present, lpxtg like motif
75	Dif 234 His	Lipoprotein signal
76	Dif 327A His	**SlpA (from SM) recombinant**
77	Dif 340 His	Leader present
78	DIF 12 His	Leader present , NLPC_P60 domain, sh3 domain (CD0183)
79	DIF 153	leader present (CD2830)
80	DIF 183 His	Lipoprotein signal (CD3669)
81	DIF192 dominio 1 His	Leader present, cw binding domains **cwp16**
82	DIF208 His	Leader present, lpxtg like motif, cna_B domains
83	DIF210 dominio1 His	Leader present, lpxtg like motif
84	DIF 251His	Leader present, lpxtg like protein (CD1858)
85	CD0855	lipoprotein
86	CD0873	lipoprotein
87	CD1653	lipoprotein
88	CD2029	lipoprotein
89	CD2672	lipoprotein

All of the selected proteins belong to the *C. difficile* strain 630, except for Dif 327A, HMW_SM and FliC-SM that belong to the strain R20291. Among the proteins selected for this screening 14 are fragments of toxins A and B, representing different domains of the two toxins.

The mature form of each *C. difficile* protein was overexpressed in *E. coli* as an N-terminal His-tag fusion protein and purified from the bacterial soluble fraction. After purification, the proteins were printed onto nitrocellulose-coated glass slides in sixteen replicates.

Serial dilutions of Biotinylated Bovine Serum Albumin (BSA), Cy3/Cy5-labeled BSA, Mouse, Human and Hamster IgG were also spotted on each array as controls for fluorophore-conjugated secondary antibodies, and for the behaviour of every pin on each printed slide during the spotting session. Proteins were spotted onto nitrocellulose-coated slides with the Arrayjet spotter.

Hamster IgGs (HaIgG) were also spotted onto the array at 5 serial dilutions (concentrations from 0.008 to 0.5 mg/ml) and used as detection and calibration controls. Validation of the experimental conditions used for detecting protein-antibody recognition was obtained by observing that the Ha-IgG values of Mean Fluorescence Intensities (MFI) of HaIgG spots, detected after exposure with Alexa Fluor647-conjugated secondary antibody, were fitted best by sigmoid curves (Figure 26).



	HaIgG 0,004 mg/ml	HaIgG 0,008 mg/ml	HaIgG 0,016 mg/ml	HaIgG 0,032 mg/ml	HaIgG 0,064 mg/ml	HaIgG 0,125 mg/ml	HaIgG 0,25 mg/ml	HaIgG 0,5 mg/ml
serum 1	80	567	2002	6000	12559	27446	57576	64823
serum 2	84	662	2473	7199	15103	33632	61499	64803
serum 3	97	658	2406	6941	14279	30797	60266	64940

Figure 26. Graphic representation of the Ha-IgG control curves. On the x axis are indicated the different Ha-IgG concentrations, MFI values are reported on the y axis and the continuous line correspond to the interpolated resulting curve. The figure is the graphic representation of the distribution of MFI values measured after incubation with Alexa Fluor647-conjugated secondary antibody.

Other samples were printed onto the array and used as controls for spotting and detection. These included PBS–Glycerol buffer, which was used to detect protein carryover during spotting and for setting the background value, and NN-His extract (cell extract of *E. coli* not expressing any *C. difficile* genes), which were used to detect aspecific signals.

On the basis of the results of the validation experiments, we arbitrarily set as a constraint for scoring a positive result as a real recognition, in a scale of MFI from 0 to 65000, a value greater than 2000 MFI.

To identify antigens recognised by the hamster sera, each serum was hybridized to the microarray and antigen recognition was detected with fluorophore-labelled secondary antibodies, Anti-hamster IgG/Alexa Fluor647 (Invitrogen, cod.A21451) or Anti-hamster

IgG/cy3 (Abcam, code: ab6969). All the incubation steps were conducted under agitation using the HS 4800 hybridization station (TECAN).

The arrays were incubated with the hamster sera reported in Table 6

Table 6. Sera obtained from hamsters in protection experiments.

serum	infecting strain	immunization	hamster	clinical outcome
1 (negative control)	------	------	Not infected	------
2	630	no vaccine	H9	died
3	630	p5/6 + ToxB-B	H1-H6	survived + diarrhea
4	630	p5/6 + ToxB-GT	H1-H6	No diarrhea
5	B1	no vaccine	H9-H10	died
6	B1	p5/6 + ToxB-B	H2-H6	survived + diarrhea
7	B1	p5/6 + ToxB-B + ToxB-GT	H2	died
8	B1	p5/6 + ToxB-B + ToxB-GT	H1,H3-H6	survived + diarrhea

There were two sera (number 2 and 5) obtained from control hamsters infected with *C. difficile* without previous vaccination. Also, when more than one hamster was used for a specific immunization/infection scheme, the sera of hamsters belonging to the same group were pooled prior to hybridization to the protein arrays.

Table 7 reports the MFI values obtained for all the protein spots. Overall, 39 samples, which include 7 different CWP proteins, showed MFI values higher than 2000 and were scored as positive. MFI values reported in Table 6 have already been normalised by subtracting the value of the negative control (NNHis).

Table 7. MFI values obtained after incubation with hamster sera.

serum number	control (no serum)	1	2	3	4	5	6	7	8
clinic			died	Survived+diarrhoea	No diarrhoea	Died	Survived+diarrhoea	died	Survived+diarrhoea
Recognized protein	control (no serum)	serum 1	serum 2	serum 3	serum 4	serum 5	serum 6	serum 7	serum 8
A16 His	16	966	334	18612	0	1400	44051	7960	25133
B4 His	70	1533	1956	63757	62597	3484	60156	57912	54158
Chim9 His	23	87	0	15	1687	163	11835	27008	44180
N-G-BB His	74	4689	1916	63246	60603	7645	60445	27588	54390
ToxA-B2 His	87	899	307	63209	62014	1140	60045	23955	38026
ToxA-GT WT	176	4035	2193	1601	3026	7474	0	907	0
ToxB-GT His	15	806	235	532	12738	1521	14221	37110	35982
Tox B1 His	16	660	230	58107	60183	1142	59035	37213	51591
Tox B4 His	95	480	143	61556	61602	790	59525	47905	53109
ToxB4	30	565	50	59105	60197	868	58768	44785	50238
FliC_014	23	153	0	15316	13978	129	11479	3251	2327
Dif2 His	40	439	562	1314	5804	1073	10118	14200	29378
Dif14B His	8	89	0	3593	0	216	5752	0	5272
Dif15 His	46	621	345	1981	7430	1345	24898	49366	50938
Dif40 His	96	692	168	2466	1207	1274	5773	451	11097
Dif51 His	53	3906	1024	3146	4700	6186	11830	3897	23887
Dif55 His	98	689	271	2290	2154	1151	7983	10186	19008
Dif75A His	30	464	626	965	10042	1337	17778	46193	53744
Dif104 His	53	1521	699	2388	2088	4587	4761	213	4457
Dif109B His	66	800	392	877	1276	1842	4813	4612	12636
Dif130 His	65	2333	1253	2512	3979	7051	5165	1041	9077
Dif170 His	0	190	0	2319	0	357	1216	0	0
Dif173 His	29	305	0	3097	1801	419	1686	0	0
Dif189 A His (CwpV)	99	1229	647	3932	6160	2753	23015	26924	54363
Dif189 B His (CwpV)	52	926	1767	4205	16659	3914	26119	53007	54837
Dif192 His (Cwp16)	49	441	426	3074	3879	1332	14226	8353	31260
Dif194 His	34	497	234	1115	504	1061	2625	2499	7663
Dif196 His (Cwp13)	0	88	36	2436	1582	319	4948	0	8191
Dif207 His (Cwp10)	88	1720	696	2139	1613	3977	12876	9073	37522
Dif208A His	19	224	231	2126	955	812	7872	547	12480
Dif208B His	5	319	326	1484	1218	968	8101	5636	21041
Dif210 His	24	367	758	2802	939	2154	5473	1067	10639
Dif211 His (Cwp21)	145	4742	2230	3476	22289	8770	40751	59778	54329
Dif212 His	21	493	416	2588	7193	1529	21886	44111	53560
Dif225 His	22	344	111	2590	1599	1076	8497	2354	13307
Dif231 His	8	2447	1015	263	388	4294	0	0	0
Dif232 His	27	954	422	331	1844	1518	15740	25476	40299
Dif234 His	0	301	458	2611	2195	789	4705	2590	9501
Dif327A His (LMW SM)	0	0	0	2192	0	0	3191	0	2227
PBS_Gly	7	13	10	364	395	18	501	115	406
HatgG 0,004 mg/ml	62	80	84	97	78	141	83	57	45
HatgG 0,008 mg/ml	452	567	662	658	607	455	571	456	339
HatgG 0,016 mg/ml	1815	2002	2473	2406	2260	1764	2260	1644	1160
HatgG 0,032 mg/ml	5578	6000	7199	6941	6666	5515	6443	4346	2882
HatgG 0,064 mg/ml	12829	12559	15103	14279	13652	12154	14336	8653	6318
HatgG 0,125 mg/ml	27158	27446	33632	30797	30949	27231	31495	18277	13389
HatgG 0,25 mg/ml	44948	57576	61499	60266	59552	58242	60176	44155	27470
HatgG 0,5 mg/ml	64352	64823	64803	64940	64738	64930	64888	64393	62209

Legend

MFI : mean fluorescence intensity

MFI>2000

MFI>5000

MFI>15000

The results obtained specifically for the toxin A and B fragments are presented in Figure 2. These data clearly show that the highest MFI values were observed with sera from vaccinated hamsters while, as expected, the sera from hamsters infected without a previous vaccination displayed very low values. The toxins that show the highest MFI values are ToxB4, ToxA-B2 His and ToxB1 (Figure 27).

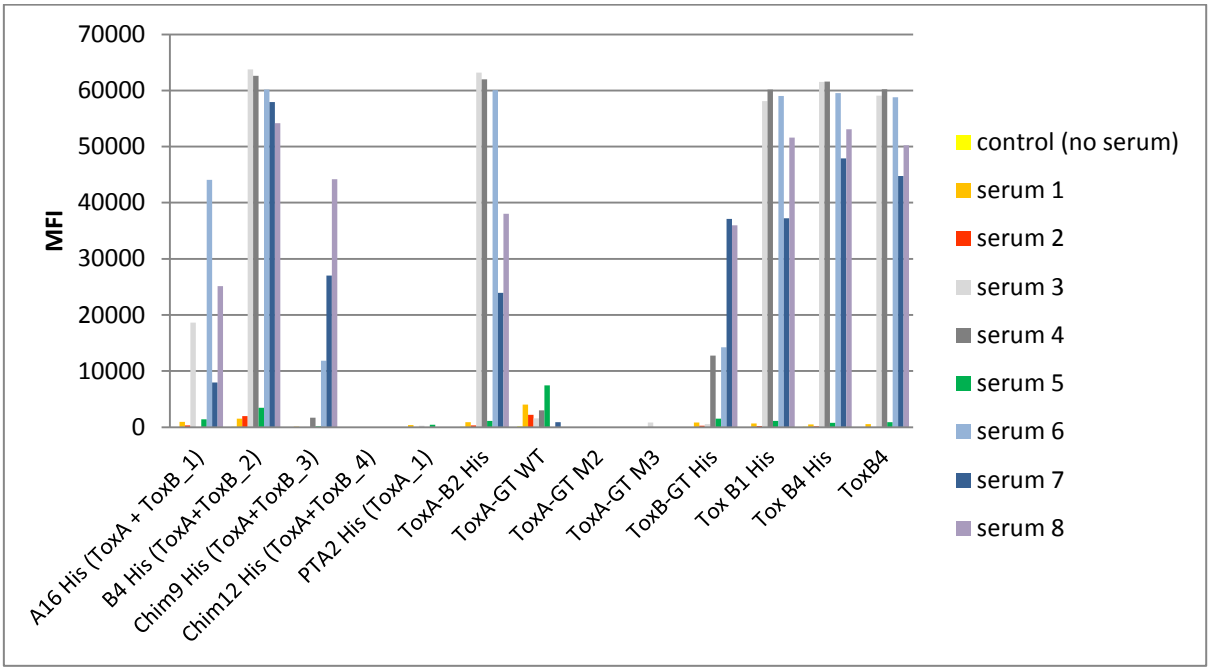


Figure 27. Protein microarray results obtained specifically for the toxin A and B fragments.

When comparing the values observed for the toxins with those of proteins not used for vaccination (Figure 28), it is noted that there was a good level of recognition by a number of sera with several surface proteins: Dif15, Dif75A, Dif212. Similar results were obtained also for three CWPs (particularly with sera 7 and 8): CwpV and Cwp21.

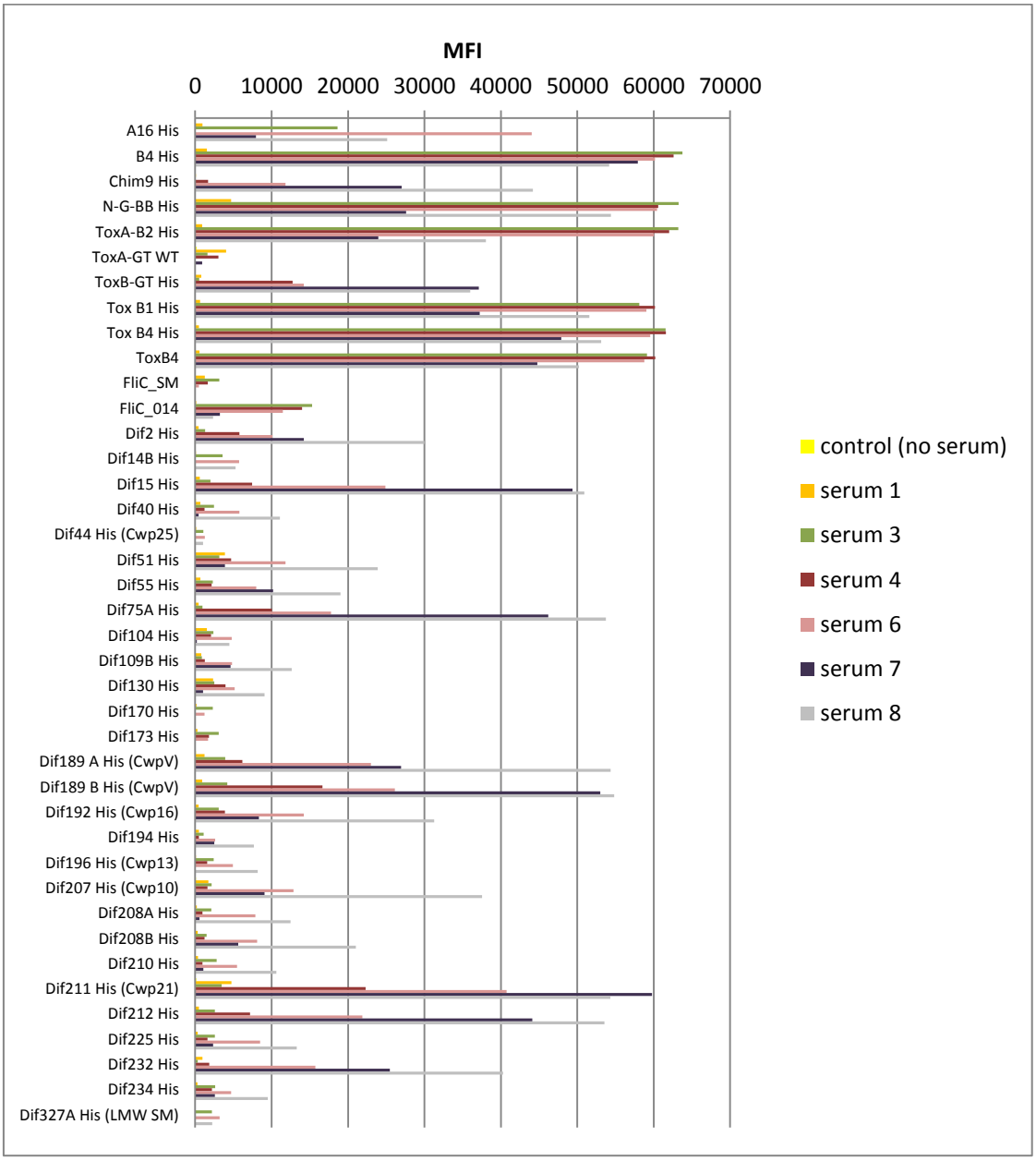


Figure 28. Comparison between values observed for the toxins with those of proteins not used for vaccination. In this figure are shown the results obtained with sera from all vaccinated hamsters.

The same proteins gave different results when using the hamster sera obtained from non-vaccinated infected hamsters (Figure 29). In fact, with these sera MFI values are very low for all the proteins. The serum that showed higher values was the serum (number 4) from the hamster that was challenged with strain B1. However, it should also be noted that the results obtained with serum 2 (hamster not vaccinated and not infected) are above our threshold (2000 MFI) for several of the spotted proteins.

Hence, we will consider as positive only values that are at least two-fold higher than those obtained for the same samples from the serum 9 “negative control”. Based on this consideration, the proteins that we can postulate as being expressed *in vivo* at levels sufficiently high to elicit a good immune response are: ToxA-GT, Dif104, Cwp21, CwpV, Cwp10.

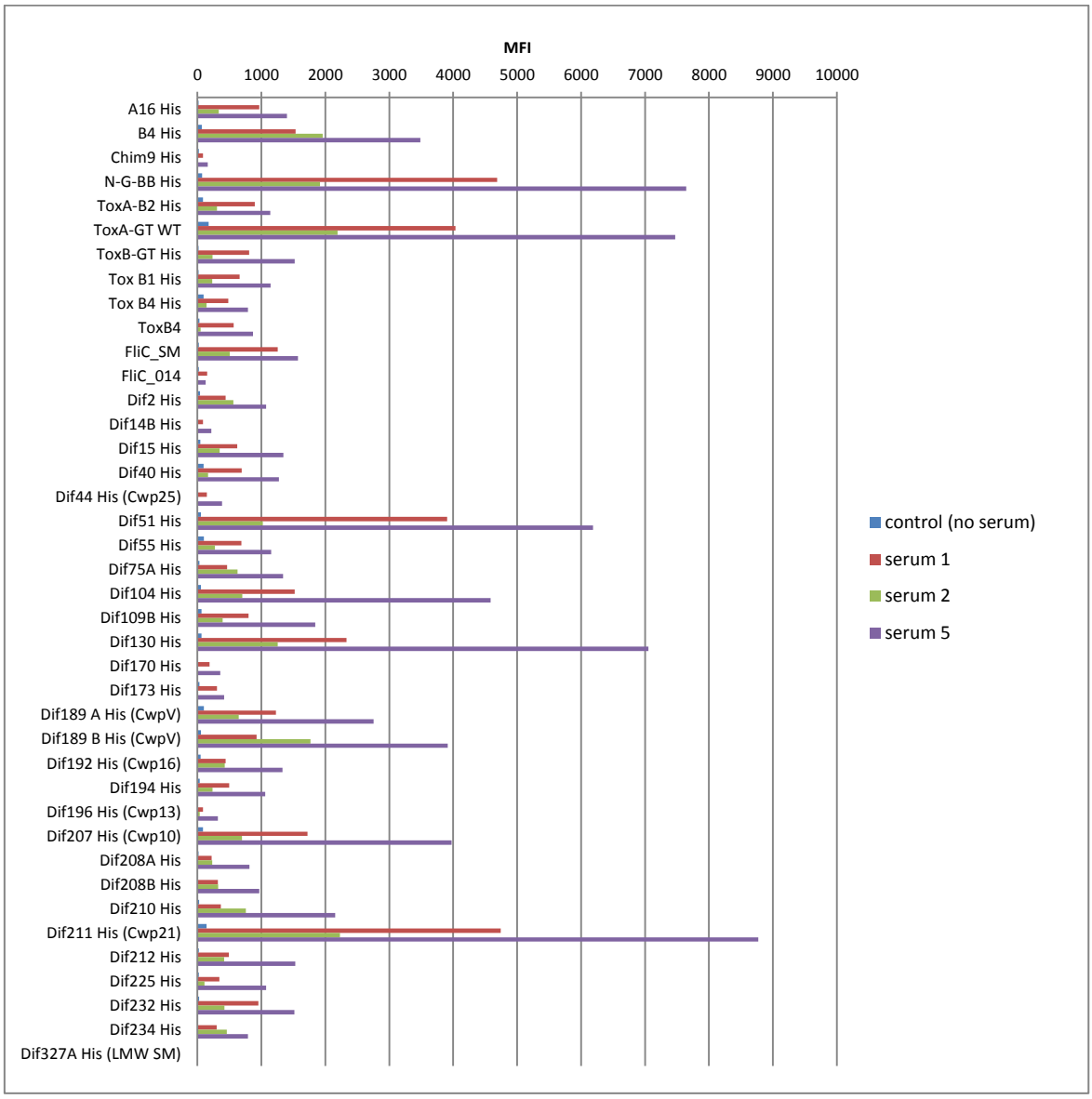


Figure 29 .Comparison between values observed for the toxins with those of proteins not used for vaccination. In this figure are shown the results obtained with sera from all non-vaccinated hamsters

To demonstrate the reproducibility of the results obtained, we repeated the hybridization experiment with the serum 3 sample changing only the method of detection. In particular, a secondary anti-hamster Cy3-conjugated antibody was used instead of the Alexa Fluor 647-labelled antibody used in previous experiments. The results obtained (Figure 30)

are similar, and the differences are possibly due to a generally less intense emission when using the Cy3 fluorophore.

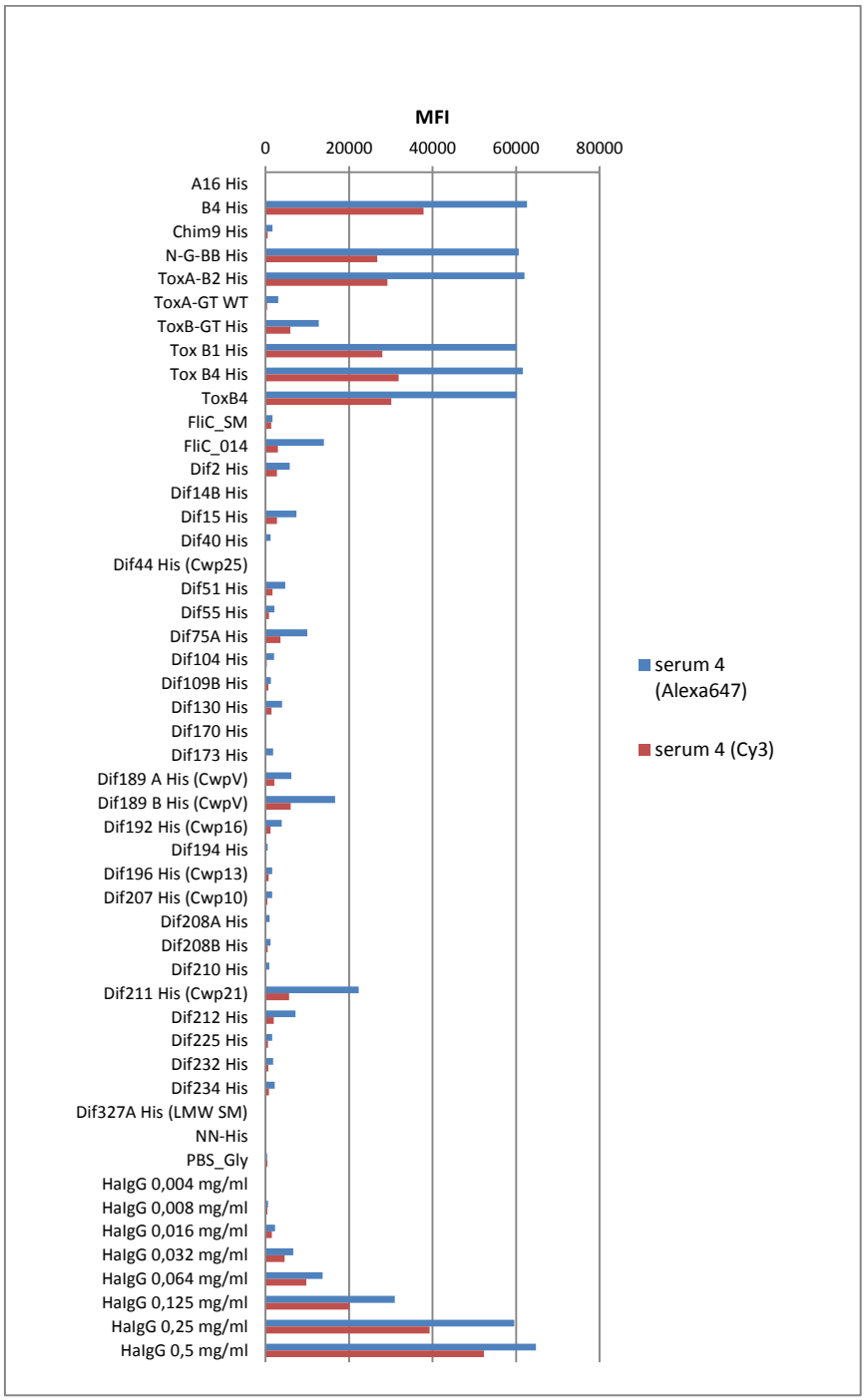


Figure 30. Comparison between values observed for serum 3 and two different secondary antibodies : Alexa Fluor 647 and Cy3

4. DISCUSSION

The observation that the emergence of new *C. difficile* strains is associated with an increased incidence and virulence of CDI suggests that strain differences play an important role in the onset and subsequent outcome of disease. For this reason, many recent studies have focused on the characterisation of the genetic variability found in clinical isolates belonging to PCR ribotypes recurrent in outbreaks of CDI (Stabler, Gerding et al. 2006; He, Sebaihia et al. 2010; Scaria, Ponnala et al. 2010; Forgetta, Oughton et al. 2011). Among the genetic traits considered to be important for pathogenicity, we have chosen to investigate the variability of a number of genes coding for a family of surface exposed proteins, the CWPs or cell wall proteins.

Our analysis was carried out on 14 of the 29 known *cwp* genes in 40 Italian clinical isolates that belong to PCR ribotypes 001, 012, 014, 018, 078 and 126. The data have provided insight into the extent of sequence variability between strains of different PCR ribotypes as well as between different isolates of the same PCR ribotype. On the basis of the degree of sequence conservation, these *cwp* genes could be divided into two groups. One comprises 9 highly conserved *cwp* genes that have identical sequences in all the isolates of the same PCR ribotype and only a few polymorphisms between PCR ribotypes, and the other group includes 5 variable *cwp* genes with low sequence conservation between isolates of the same PCR ribotype as well as between different PCR ribotypes. Interestingly, the latter group comprises *cwp27* and *cwp29*, two genes coding for CWPs that do not contain putative domains assigned to a known function (Fagan and Fairweather 2011). A search for sequence homology of their unassigned C-terminal regions, however, showed that both have some sequence similarity with phage proteins (data not shown); thus implying that these genes could have been acquired through horizontal gene transfer events. All the other Cwps contain two or three cell wall binding motifs in addition to a unique domain that is proposed to specify a function.

Moreover, the results of our phylogenetic analysis show that certain PCR ribotypes always display the same type of variability for most of the *cwp* genes included in our study. This is the case for PCR ribotypes 014 and 018, or 078 and 126. In particular, all ten PCR ribotype 078 isolates of our collection have polymorphisms identical to those found in the six PCR ribotype 126 strains, while they differ, albeit just for a few nucleotides, from the 078 reference strains M120 and QCD23M63 (Figure 22).

This is in agreement with previous studies, where PCR ribotypes 078 and 126 are always assigned to the same lineage (Spigaglia, Barbanti et al. 2010; Reil, Erhard et al. 2011),

a clade which is frequently associated with livestock as well as with humans (Keel, Brazier et al. 2007; Goorhuis, Debast et al. 2008; Hensgens, Keessen et al. 2012). Another interesting outcome of our study is the finding that, with regard to *cwp* gene variability, PCR ribotypes 078/126 are more closely related to PCR ribotype 027 than to any of the other PCR ribotypes analysed (Figure 22). Recently, a number of reports have focused their interest on determining which genetic factors PCR ribotypes 027 and 078/126 may have in common that would help to explain the similarity in CDI outcome observed for these "hypervirulent" strains (Knetsch, Hensgens et al. 2011; Barbut and Rupnik 2012; Walk, Micic et al. 2012). Our data clearly suggest that 14 of the 29 predicted CWPs share a high degree of sequence similarity in strains that belong to PCR ribotypes 078/126 and 027. As CWPs are surface components of *C. difficile* possibly involved in colonisation and onset of CDI, it is proposed that several of the 14 CWPs characterised in this work may represent common traits of these PCR ribotypes which contribute to their "hypervirulent" behaviour.

The genetic diversity found among the six PCR ribotypes is not evenly distributed between the 14 *cwp* genes of interest and can be used to discriminate between highly conserved and variable *cwp* genes. Likewise, expression of conserved CWPs seems to be well conserved in all isolates (Figure 23), while variable CWPs are not always expressed at comparable levels even in strains containing identical sequences but belonging to different PCR ribotypes, as seen for the expression of Cwp27 in PCR ribotype 001 isolates (Figure 25). Our results highlight how difficult it is to characterise key components of the *C. difficile* cell surface due to the exceedingly high overall genetic complexity present in different *C. difficile* isolates.

One approach to identify the proteins that could be immunogenic is to examine the hamster immune response to bacterial infection. This not only provides information about the expression of proteins in the host, but also may serve to identify proteins involved in pathogenesis. In fact, by the use of protein microarray analysis, some Cwps such as Cwp10, CwpV and Cwp21, were shown to be the most immunogenic Cwps under our conditions. The majority of the Cwps present on the chip are recombinant proteins, but estimates on their expression levels and *in vivo* exposure are not currently available.

For these two reasons, the low values obtained for the other Cwps should not be interpreted as lack of expression of these proteins *in vivo*, but they are more easily correlated with low immunogenic properties of these Cwps and our experimental approach, which involves the use of recombinant proteins.

The *in vitro* expression of these proteins was confirmed by our Western blot results. In the case of conserved *cwp* genes, the results obtained are very clear. Indeed, in the case of the

extremely variable *cwp26* gene, although we were not able to amplify and sequence this ortholog in several isolates (Figure 21), expression of a Cwp26-like protein could be detected in most of the PCR negative strains. Moreover, in all the PCR ribotype 001, 012, 014 and 018 strains analysed expression of Cwp26 was observed only in S-layer extracts, indicating that Cwp26 is a complex constituent of the S layer in these PCR ribotypes (Figure 24B). The 70 kDa bands may indicate a propensity of Cwp26 to form stable complexes with other Cwps. As it is known that the few Cwps characterized so far are processed (de la Riva, Willing et al. 2011; Reynolds, Emerson et al. 2011), it cannot be excluded that a similar mechanism also applies to the mature form of Cwp26.

In summary, we propose that the conserved CWPs may correspond to essential components of the bacterial surface, while the highly variable CWPs could be more recent acquisitions of additional surface elements.

As the specific function of the majority of the CWPs analysed in our study remains unclear, it is not currently possible to elucidate if there is a correlation between the presence of a particular Cwp in the S-layer and *C. difficile* interspecies transmission, or increased spread and severity of CDI. All these aspects need to be addressed urgently in order to be able to contain the significant increase in disease incidence and mortality reported in recent years. Finally, due to the complexity of the genetic variability observed in *C. difficile* strains, a unique method for typing newly emerging strains of *C. difficile* is still not available. The analysis of *cwp* gene diversity could offer an additional tool for the classification of *C. difficile* clinical isolates.

5. REFERENCES

Angenendt, P., J. Glokler, et al. (2002). "Toward optimized antibody microarrays: a comparison of current microarray support materials." Anal Biochem**309**(2): 253-260.

Arcus, V. L., J. S. Lott, et al. (2006). "The potential impact of structural genomics on tuberculosis drug discovery." Drug Discov Today**11**(1-2): 28-34.

Bacci, S., G. St-Martin, et al. (2009). "Outbreak of Clostridium difficile 027 in North Zealand, Denmark, 2008-2009." Euro Surveill**14**(16).

Baldassarri, L., G. Donelli, et al. (1991). "Capsule-like structures in Clostridium difficile strains." Microbiologica**14**(4): 295-300.

Barbut, F., B. Gariazzo, et al. (2007). "Clinical features of Clostridium difficile-associated infections and molecular characterization of strains: results of a retrospective study, 2000-2004." Infect Control Hosp Epidemiol**28**(2): 131-139.

Barbut, F. and M. Rupnik (2012). "Editorial commentary: 027, 078, and others: going beyond the numbers (and away from the hypervirulence)." Clin Infect Dis**55**(12): 1669-1672.

Bartlett, J. G. (2008). "Historical perspectives on studies of Clostridium difficile and C. difficile infection." Clin Infect Dis**46 Suppl 1**: S4-11.

Bauer, M. P., D. W. Notermans, et al. (2011). "Clostridium difficile infection in Europe: a hospital-based survey." Lancet**377**(9759): 63-73.

Belyi, Y. and K. Aktories (2010). "Bacterial toxin and effector glycosyltransferases." Biochim Biophys Acta**1800**(2): 134-143.

Bensing, B. A. and P. M. Sullam (2002). "An accessory sec locus of Streptococcus gordonii is required for export of the surface protein GspB and for normal levels of binding to human platelets." Mol Microbiol**44**(4): 1081-1094.

Bianco, M., G. Fedele, et al. (2011). "Immunomodulatory activities of surface-layer proteins obtained from epidemic and hypervirulent Clostridium difficile strains." J Med Microbiol**60**(Pt 8): 1162-1167.

Borriello, S. P. (1998). "Pathogenesis of Clostridium difficile infection." J Antimicrob Chemother**41 Suppl C**: 13-19.

Borriello, S. P., H. A. Davies, et al. (1990). "Virulence factors of Clostridium difficile." Rev Infect Dis**12 Suppl 2**: S185-191.

Cairns, M. D., R. A. Stabler, et al. (2012). "The continually evolving Clostridium difficile species." Future Microbiol**7**(8): 945-957.

Calabi, E. and N. Fairweather (2002). "Patterns of sequence conservation in the S-layer proteins and related sequences in Clostridium difficile." J Bacteriol**184**(14): 3886-3897.

Calabi, E., S. Ward, et al. (2001). "Molecular characterization of the surface layer proteins from Clostridium difficile." Mol Microbiol**40**(5): 1187-1199.

Chapeton Montes, D., A. Collignon, et al. (2013). "Influence of environmental conditions on the expression and the maturation process of the Clostridium difficile surface associated protease Cwp84." Anaerobe**19**: 79-82.

Charles, P. T., E. R. Goldman, et al. (2004). "Fabrication and characterization of 3D hydrogel microarrays to measure antigenicity and antibody functionality for biosensor applications." Biosens Bioelectron**20**(4): 753-764.

Chen, Q., H. Wu, et al. (2004). "Investigating the role of secA2 in secretion and glycosylation of a fimbrial adhesin in Streptococcus parasanguis FW213." Mol Microbiol**53**(3): 843-856.

Cheng, V. C., W. C. Yam, et al. (2009). "Clostridium difficile ribotype 027 arrives in Hong Kong." Int J Antimicrob Agents**34**(5): 492-493.

Chenna, R., H. Sugawara, et al. (2003). "Multiple sequence alignment with the Clustal series of programs." Nucleic Acids Res**31**(13): 3497-3500.

Davies, H. A. and S. P. Borriello (1990). "Detection of capsule in strains of Clostridium difficile of varying virulence and toxigenicity." Microb Pathog**9**(2): 141-146.

Dawson, L. F., E. H. Donahue, et al. (2011). "The analysis of para-cresol production and tolerance in Clostridium difficile 027 and 012 strains." BMC Microbiol**11**: 86.

De Groot, A. S. (2006). "Immunomics: discovering new targets for vaccines and therapeutics." Drug Discov Today**11**(5-6): 203-209.

de la Riva, L., S. E. Willing, et al. (2011). "Roles of cysteine proteases Cwp84 and Cwp13 in biogenesis of the cell wall of Clostridium difficile." J Bacteriol**193**(13): 3276-3285.

Debast, S. B., L. A. van Leengoed, et al. (2009). "Clostridium difficile PCR ribotype 078 toxinotype V found in diarrhoeal pigs identical to isolates from affected humans." Environ Microbiol**11**(2): 505-511.

Delehanty, J. B. (2004). "Printing functional protein microarrays using piezoelectric capillaries." Methods Mol Biol**264**: 135-143.

Driessen, A. J. and N. Nouwen (2008). "Protein translocation across the bacterial cytoplasmic membrane." Annu Rev Biochem**77**: 643-667.

Dupuy, B., R. Govind, et al. (2008). "Clostridium difficile toxin synthesis is negatively regulated by TcdC." J Med Microbiol**57**(Pt 6): 685-689.

Eidhin, D. N., A. W. Ryan, et al. (2006). "Sequence and phylogenetic analysis of the gene for surface layer protein, slpA, from 14 PCR ribotypes of Clostridium difficile." J Med Microbiol**55**(Pt 1): 69-83.

Emerson, J. E., C. B. Reynolds, et al. (2009). "A novel genetic switch controls phase variable expression of CwpV, a Clostridium difficile cell wall protein." Mol Microbiol**74**(3): 541-556.

Eveillard, M., V. Fourel, et al. (1993). "Identification and characterization of adhesive factors of Clostridium difficile involved in adhesion to human colonic enterocyte-like Caco-2 and mucus-secreting HT29 cells in culture." Mol Microbiol**7**(3): 371-381.

Fagan, R. and N. Fairweather (2010). "Dissecting the cell surface." Methods Mol Biol**646**: 117-134.

Fagan, R. P., D. Albesa-Jove, et al. (2009). "Structural insights into the molecular organization of the S-layer from Clostridium difficile." Mol Microbiol**71**(5): 1308-1322.

Fagan, R. P. and N. F. Fairweather (2011). "Clostridium difficile has two parallel and essential Sec secretion systems." J Biol Chem**286**(31): 27483-27493.

Fagan, R. P., C. Janoir, et al. (2011). "A proposed nomenclature for cell wall proteins of Clostridium difficile." J Med Microbiol**60**(Pt 8): 1225-1228.

Forgetta, V., M. T. Oughton, et al. (2011). "Fourteen-genome comparison identifies DNA markers for severe-disease-associated strains of Clostridium difficile." J Clin Microbiol**49**(6): 2230-2238.

Ganeshapillai, J., E. Vinogradov, et al. (2008). "Clostridium difficile cell-surface polysaccharides composed of pentaglycosyl and hexaglycosyl phosphate repeating units." Carbohydr Res**343**(4): 703-710.

Gavin, A. C., P. Aloy, et al. (2006). "Proteome survey reveals modularity of the yeast cell machinery." Nature**440**(7084): 631-636.

Gavin, A. C., M. Bosche, et al. (2002). "Functional organization of the yeast proteome by systematic analysis of protein complexes." Nature**415**(6868): 141-147.

Ginaldi, L., M. F. Loreto, et al. (2001). "Immunosenescence and infectious diseases." Microbes Infect**3**(10): 851-857.

Goncalves, C., D. Decre, et al. (2004). "Prevalence and characterization of a binary toxin (actin-specific ADP-ribosyltransferase) from Clostridium difficile." J Clin Microbiol**42**(5): 1933-1939.

Goorhuis, A., S. B. Debast, et al. (2008). "Clostridium difficile PCR ribotype 078: an emerging strain in humans and in pigs?" J Clin Microbiol**46**(3): 1157; author reply 1158.

Gould, C. V. and L. C. McDonald (2008). "Bench-to-bedside review: Clostridium difficile colitis." Crit Care**12**(1): 203.

Hall, D. A., J. Ptacek, et al. (2007). "Protein microarray technology." Mech Ageing Dev**128**(1): 161-167.

Hall, D. A., H. Zhu, et al. (2004). "Regulation of gene expression by a metabolic enzyme." Science**306**(5695): 482-484.

He, M., M. Sebaihia, et al. (2010). "Evolutionary dynamics of Clostridium difficile over short and long time scales." Proc Natl Acad Sci U S A**107**(16): 7527-7532.

Hensgens, M. P., E. C. Keessen, et al. (2012). "Clostridium difficile infection in the community: a zoonotic disease?" Clin Microbiol Infect**18**(7): 635-645.

Ho, Y., A. Gruhler, et al. (2002). "Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry." Nature**415**(6868): 180-183.

Hookman, P. and J. S. Barkin (2009). "Clostridium difficile associated infection, diarrhea and colitis." World J Gastroenterol**15**(13): 1554-1580.

Huang, J., H. Zhu, et al. (2004). "Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips." Proc Natl Acad Sci U S A**101**(47): 16594-16599.

Johri, A. K., L. C. Paoletti, et al. (2006). "Group B Streptococcus: global incidence and vaccine development." Nat Rev Microbiol**4**(12): 932-942.

Jones, R. B., A. Gordus, et al. (2006). "A quantitative protein interaction network for the ErbB receptors using protein microarrays." Nature**439**(7073): 168-174.

Karjalainen, T., N. Saumier, et al. (2002). "Clostridium difficile genotyping based on slpA variable region in S-layer gene sequence: an alternative to serotyping." J Clin Microbiol**40**(7): 2452-2458.

Karjalainen, T., A. J. Waligora-Dupriet, et al. (2001). "Molecular and genomic analysis of genes encoding surface-anchored proteins from Clostridium difficile." Infect Immun**69**(5): 3442-3446.

Kato, H., Y. Ito, et al. (2007). "First isolation of Clostridium difficile 027 in Japan." Euro Surveill**12**(1): E070111 070113.

Kattah, M. G., G. R. Alemi, et al. (2006). "A new two-color Fab labeling method for autoantigen protein microarrays." Nat Methods**3**(9): 745-751.

Keel, K., J. S. Brazier, et al. (2007). "Prevalence of PCR ribotypes among Clostridium difficile isolates from pigs, calves, and other species." J Clin Microbiol**45**(6): 1963-1964.

Khanna, S., D. S. Pardi, et al. (2012). "The epidemiology of community-acquired Clostridium difficile infection: a population-based study." Am J Gastroenterol**107**(1): 89-95.

Killgore, G., A. Thompson, et al. (2008). "Comparison of seven techniques for typing international epidemic strains of Clostridium difficile: restriction endonuclease analysis, pulsed-field gel electrophoresis, PCR-ribotyping, multilocus sequence typing, multilocus variable-number tandem-repeat analysis, amplified fragment length polymorphism, and surface layer protein A gene sequence typing." J Clin Microbiol**46**(2): 431-437.

Kirby, J. M., H. Ahern, et al. (2009). "Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of Clostridium difficile." J Biol Chem**284**(50): 34666-34673.

Klock, H. E., E. J. Koesema, et al. (2008). "Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts." Proteins**71**(2): 982-994.

Klock, H. E., A. White, et al. (2005). "Methods and results for semi-automated cloning using integrated robotics." J Struct Funct Genomics**6**(2-3): 89-94.

Knetsch, C. W., M. P. Hensgens, et al. (2011). "Genetic markers for Clostridium difficile lineages linked to hypervirulence." Microbiology**157**(Pt 11): 3113-3123.

Kuijper, E. J., F. Barbut, et al. (2008). "Update of Clostridium difficile infection due to PCR ribotype 027 in Europe, 2008." Euro Surveill**13**(31).

Kusnezow, W., A. Jacob, et al. (2003). "Antibody microarrays: an evaluation of production parameters." Proteomics**3**(3): 254-264.

Lawley, T. D., N. J. Croucher, et al. (2009). "Proteomic and genomic characterization of highly infectious Clostridium difficile 630 spores." J Bacteriol**191**(17): 5377-5386.

Lenz, L. L. and D. A. Portnoy (2002). "Identification of a second Listeria secA gene associated with protein secretion and the rough phenotype." Mol Microbiol**45**(4): 1043-1056.

Lo Vecchio, A. and G. M. Zacur (2012). "Clostridium difficile infection: an update on epidemiology, risk factors, and therapeutic options." Curr Opin Gastroenterol**28**(1): 1-9.

Loo, V. G., L. Poirier, et al. (2005). "A predominantly clonal multi-institutional outbreak of Clostridium difficile-associated diarrhea with high morbidity and mortality." N Engl J Med**353**(23): 2442-2449.

Louie, T. J., M. A. Miller, et al. (2011). "Fidaxomicin versus vancomycin for Clostridium difficile infection." N Engl J Med**364**(5): 422-431.

MacBeath, G. and S. L. Schreiber (2000). "Printing proteins as microarrays for high-throughput function determination." Science**289**(5485): 1760-1763.

McCoubrey, J. and I. R. Poxton (2001). "Variation in the surface layer proteins of Clostridium difficile." FEMS Immunol Med Microbiol**31**(2): 131-135.

McCusker, M. E., A. D. Harris, et al. (2003). "Fluoroquinolone use and Clostridium difficile-associated diarrhea." Emerg Infect Dis**9**(6): 730-733.

McDonald, L. C., G. E. Killgore, et al. (2005). "An epidemic, toxin gene-variant strain of Clostridium difficile." N Engl J Med**353**(23): 2433-2441.

Mendoza, L. G., P. McQuary, et al. (1999). "High-throughput microarray-based enzyme-linked immunosorbent assay (ELISA)." Biotechniques**27**(4): 778-780, 782-776, 788.

Metcalf, D., B. P. Avery, et al. (2011). "Clostridium difficile in seafood and fish." Anaerobe**17**(2): 85-86.

Michaud, G. A., M. Salcius, et al. (2003). "Analyzing antibody specificity with whole proteome microarrays." Nat Biotechnol**21**(12): 1509-1512.

Mora, M., C. Donati, et al. (2006). "Microbial genomes and vaccine design: refinements to the classical reverse vaccinology approach." Curr Opin Microbiol**9**(5): 532-536.

Mora, M., D. Veggi, et al. (2003). "Reverse vaccinology." Drug Discov Today**8**(10): 459-464.

Nakai, K. and P. Horton (1999). "PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization." Trends Biochem Sci**24**(1): 34-36.

Owens, R. C. (2007). "Clostridium difficile-associated disease: changing epidemiology and implications for management." Drugs**67**(4): 487-502.

Papatheodorou, P., C. Wilczek, et al. (2012). "Identification of the cellular receptor of Clostridium spiroforme toxin." Infect Immun**80**(4): 1418-1423.

Pepin, J., L. Valiquette, et al. (2004). "Clostridium difficile-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity." CMAJ**171**(5): 466-472.

Popoff, M. R., E. J. Rubin, et al. (1988). "Actin-specific ADP-ribosyltransferase produced by a Clostridium difficile strain." Infect Immun**56**(9): 2299-2306.

Poxton, I. R. and T. D. Cartmill (1982). "Immunochemistry of the cell-surface carbohydrate antigens of Clostridium difficile." J Gen Microbiol**128**(6): 1365-1370.

Ptacek, J., G. Devgan, et al. (2005). "Global analysis of protein phosphorylation in yeast." Nature**438**(7068): 679-684.

Quesada-Gomez, C., C. Rodriguez, et al. (2010). "Emergence of Clostridium difficile NAP1 in Latin America." J Clin Microbiol**48**(2): 669-670.

Rappuoli, R. (2000). "Reverse vaccinology." Curr Opin Microbiol**3**(5): 445-450.

Reil, M., M. Erhard, et al. (2011). "Recognition of Clostridium difficile PCR ribotypes 001, 027 and 126/078 using an extended MALDI-TOF MS system." Eur J Clin Microbiol Infect Dis**30**(11): 1431-1436.

Reynolds, C. B., J. E. Emerson, et al. (2011). "The Clostridium difficile cell wall protein CwpV is antigenically variable between strains, but exhibits conserved aggregation-promoting function." PLoS Pathog**7**(4): e1002024.

Riley, T. V., S. Thean, et al. (2009). "First Australian isolation of epidemic Clostridium difficile PCR ribotype 027." Med J Aust**190**(12): 706-708.

Rodriguez-Palacios, A., R. J. Reid-Smith, et al. (2009). "Possible seasonality of Clostridium difficile in retail meat, Canada." Emerg Infect Dis**15**(5): 802-805.

Rupnik, M., A. Widmer, et al. (2008). "Clostridium difficile toxinotype V, ribotype 078, in animals and humans." J Clin Microbiol**46**(6): 2146.

Rupnik, M., M. H. Wilcox, et al. (2009). "Clostridium difficile infection: new developments in epidemiology and pathogenesis." Nat Rev Microbiol**7**(7): 526-536.

Scaria, J., L. Ponnala, et al. (2010). "Analysis of ultra low genome conservation in Clostridium difficile." PLoS One**5**(12): e15147.

Schena, M., D. Shalon, et al. (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray." Science**270**(5235): 467-470.

Schroeder, M. S. (2005). "Clostridium difficile--associated diarrhea." Am Fam Physician**71**(5): 921-928.

Schwan, C., B. Stecher, et al. (2009). "Clostridium difficile toxin CDT induces formation of microtubule-based protrusions and increases adherence of bacteria." PLoS Pathog**5**(10): e1000626.

Schweitzer, B., S. Wiltshire, et al. (2000). "Immunoassays with rolling circle DNA amplification: a versatile platform for ultrasensitive antigen detection." Proc Natl Acad Sci U S A**97**(18): 10113-10119.

Sebaihia, M., B. W. Wren, et al. (2006). "The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome." Nat Genet**38**(7): 779-786.

Siboo, I. R., D. O. Chaffin, et al. (2008). "Characterization of the accessory Sec system of Staphylococcus aureus." J Bacteriol**190**(18): 6188-6196.

Simango, C. and S. Mwakurudza (2008). "Clostridium difficile in broiler chickens sold at market places in Zimbabwe and their antimicrobial susceptibility." Int J Food Microbiol**124**(3): 268-270.

Songer, J. G., H. T. Trinh, et al. (2009). "Clostridium difficile in retail meat products, USA, 2007." Emerg Infect Dis**15**(5): 819-821.

Spigaglia, P., F. Barbanti, et al. (2010). "Clostridium difficile isolates resistant to fluoroquinolones in Italy: emergence of PCR ribotype 018." J Clin Microbiol**48**(8): 2892-2896.

Stabler, R. A., D. N. Gerding, et al. (2006). "Comparative phylogenomics of Clostridium difficile reveals clade specificity and microevolution of hypervirulent strains." J Bacteriol**188**(20): 7297-7305.

Tae, C. H., S. A. Jung, et al. (2009). "The first case of antibiotic-associated colitis by Clostridium difficile PCR ribotype 027 in Korea." J Korean Med Sci**24**(3): 520-524.

Templin, M. F., D. Stoll, et al. (2002). "Protein microarray technology." Trends Biotechnol**20**(4): 160-166.

van Steenbergen, J., S. Debast, et al. (2005). "Isolation of Clostridium difficile ribotype 027, toxinotype III in the Netherlands after increase in C. difficile-associated diarrhoea." Euro Surveill**10**(7): E050714 050711.

von Eichel-Streiber, C., P. Boquet, et al. (1996). "Large clostridial cytotoxins--a family of glycosyltransferases modifying small GTP-binding proteins." Trends Microbiol**4**(10): 375-382.

Voth, D. E. and J. D. Ballard (2005). "Clostridium difficile toxins: mechanism of action and role in disease." Clin Microbiol Rev**18**(2): 247-263.

Walk, S. T., D. Micic, et al. (2012). "Clostridium difficile ribotype does not predict severe infection." Clin Infect Dis**55**(12): 1661-1668.

Walters, B. A., R. Roberts, et al. (1983). "Relapse of antibiotic associated colitis: endogenous persistence of Clostridium difficile during vancomycin therapy." Gut**24**(3): 206-212.

Winssinger, N., S. Ficarro, et al. (2002). "Profiling protein function with small molecule microarrays." Proc Natl Acad Sci U S A**99**(17): 11139-11144.

Wright, A., D. Drudy, et al. (2008). "Immunoreactive cell wall proteins of Clostridium difficile identified by human sera." J Med Microbiol**57**(Pt 6): 750-756.

Zhu, H., M. Bilgin, et al. (2001). "Global analysis of protein activities using proteome chips." Science**293**(5537): 2101-2105.

Zhu, H., M. Bilgin, et al. (2003). "Proteomics." Annu Rev Biochem**72**: 783-812.

Zhu, H., S. Hu, et al. (2006). "Severe acute respiratory syndrome diagnostics using a coronavirus protein microarray." Proc Natl Acad Sci U S A**103**(11): 4011-4016.