Analysis of the Amino Acid Sequence Variation of the 67–72p Protein and the Structural Pili Proteins of Corynebacterium diphtheriae for their Suitability as Potential Vaccine Antigens

KLAUDIA BRODZIK1, KATARZYNA KRYSZTOPA-GRZYBOWSKA1, MACIEJ POLAK1, JAKUB LACH2, DOMINIK STRAPAGIEL2,3 and ALEKSANDRA ANNA ZASADA1

1 Department of Sera and Vaccines Evaluation, National Institute of Public Health – National Institute of Hygiene, Warsaw, Poland
2 Biobank Lab, Department of Molecular Biophysics, Faculty of Biology and Environmental Protection, University of Lodz, Lodz, Poland
3 BBMRI.pl Consortium, Wroclaw Research Center EIT+, Wroclaw, Poland

Submitted 4 December 2018, revised 8 March 2019, accepted 9 March 2019

Abstract

The aim of this study was to identify the potential vaccine antigens in Corynebacterium diphtheriae strains by in silico analysis of the amino acid variation in the 67–72p surface protein that is involved in the colonization and induction of epithelial cell apoptosis in the early stages of infection. The analysis of pili structural proteins involved in bacterial adherence to host cells and related to various types of infections was also performed. A polymerase chain reaction (PCR) was carried out to amplify the genes encoding the 67–72p protein and three pili structural proteins (SpaC, SpaI, SapD) and the products obtained were sequenced. The nucleotide sequences of the particular genes were translated into amino acid sequences, which were then matched among all the tested strains using bioinformatics tools. In the last step, the affinity of the tested proteins to major histocompatibility complex (MHC) classes I and II, and linear B-cell epitopes was analyzed. The variations in the nucleotide sequence of the 67–72p protein and pili structural proteins among C. diphtheriae strains isolated from various infections were noted. A transposition of the insertion sequence within the gene encoding the SpaC pili structural proteins was also detected. In addition, the bioinformatics analyses enabled the identification of epitopes for B-cells and T-cells in the conserved regions of the proteins, thus, demonstrating that these proteins could be used as antigens in the potential vaccine development. The results identified the most conserved regions in all tested proteins that are exposed on the surface of C. diphtheriae cells.

Key words: Corynebacterium diphtheriae, non-toxigenic, pili, 67–72p protein, vaccine

Introduction

Corynebacterium diphtheriae is the etiologic agent of a serious infectious disease – diphtheria. The diphtheria vaccine is highly effective but is directed only against the diphtheria toxin. The non-toxigenic C. diphtheriae strains may cause many severe invasive diseases, e.g., endocarditis, septic arthritis, bacteremia, and noninvasive wound infections. The notified increasing number of non-toxigenic C. diphtheriae infections indicates that infections are a growing problem in Europe (Belko et al. 2000; Zasada et al. 2010; Zasada 2013; Frichchione et al. 2014; Dangel et al. 2018). It is hypothesized that high vaccination coverage has resulted in the emergence of non-toxin-producing (non-toxigenic) C. diphtheriae strains, which acquired new virulence factors.

The diphtheria toxoid vaccination protects against the action of the toxin but does not protect against colonization and invasion by C. diphtheriae. Little is known about C. diphtheriae virulence factors other than the diphtheria toxin. The degree of adhesion of microorganisms to host cells has been shown to be an important pathogenicity factor for both toxigenic and non-toxigenic strains (Colombo et al. 2001). The virulence factors that facilitate bacterial colonization to specific host tissues and are related to pathogenesis include pili and fimbriae (Reardon-Robinson and Ton-That 2014). These structures are covalently attached to
the bacterial cell wall and are recognized by the related host receptors (Sauer et al. 2000; Rogers et al. 2011). The surface structures are potential candidates for the development of new vaccines and antimicrobial therapies, due to their significant role in pathogenesis (Maione et al. 2005; Soriani and Telford 2010). Pili are found in both Gram-positive and Gram-negative bacteria, albeit with different folding mechanisms (Thanassi et al. 1998; Ton-That and Schneewind 2004). The function of pili is not only the involvement in adhesion, but they also act as bacteriophage receptors and participate in DNA transfer, biofilm formation, cell aggregation, host cell penetration, and motility (Proff and Baker 2009).

There are three types of pili in *C. diphtheriae* strains (SpaA, SapD, and SpaH), each containing the LPXTG motif (Ton-That and Schneewind 2003; Gaspar and Ton-That 2006; Swierczynski and Ton-That 2006). The genes involved in the production of pili encode nine pili proteins, defined in the successive letters from SpaA to Spal, and five sortases defined from SrtA to SrtE, which are organized in three separate clusters. The sixth sortase SrtF (class-D homolog) is now referred to as the housekeeping sortase, located in a different region of the chromosome (Ton-That and Schneewind 2003). *C. diphtheriae*, like other Gram-positive bacteria (e.g. *Streptococcus pneumoniae*, group A and B streptococci or *Actinomyces*), have the gene encoding cysteine transpeptidase (sortase) conserved in the genome, which is necessary for the assembly of pili (Ton-That and Schneewind 2004; Marraffini et al. 2006).

Pili are composed of three proteins: the main subunit forming the stem of pili, and two smaller subunits located at the base and at the end of the pili, e.g., SpaA-type pili is structured in such way that the SpaA pili protein creates a stem, SpaC is located at the end of pilus, while SpaB is located along the stem and at the base (Ton-That and Schneewind 2003; Mandlik et al. 2008; Rogers et al. 2011). SpaA is important for the formation of the pili structure (Ton-That and Schneewind 2004). It has been proven that in the absence of SpaA protein, SpaB and SpaC are anchored in the cell wall as monomers (Mandlik et al. 2007).

The adhesion process of *C. diphtheriae* strains to the surface of human cells is multifactorial. Functions and mechanisms of action of fimbriae (Mandlik et al. 2007), non-fimbrial 67–72p adhesin (Colombo et al. 2001), trans-sialidase (Mattos-Guaraldi et al. 1998), hydrophobins, and sugar residues (Mattos-Guaraldi et al. 1999a; 2000; Moreira et al. 2003) are poorly understood, especially how they jointly participate in the adherence to the host cells and in the colonization of these cells during bacterial infection.

Initially, the 67–72p adhesive protein was described as a ligand responsible for the adherence of *C. diphtheriae* to human erythrocytes (Colombo et al. 2001). Later, the participation of this protein in the adherence of bacteria to HEp-2 cells was described (Hirata et al. 2004). The presence of the 67–72p protein has been confirmed in *C. diphtheriae* strains isolated from various sources, e.g., on the surface of the cells of invasive HC01 strain isolated from the blood of a patient with endocarditis (Sabbadini et al. 2012).

*C. diphtheriae* strains exhibit cell surface hydrophobicity and autoaggregation. Thanks to these features, microorganisms avoid immune defenses and are able to survive on the surface of the skin and mucosal membranes (Mattos-Guaraldi et al. 1999b). In addition, it has been proven that the 67–72p protein has the ability to induce host cell death, giving a signal for apoptosis in the early stages of infection. The occurrence of 67–72p hemagglutinin is one of the characteristics of the potentially invasive strains because it can contribute to the cytotoxicity and apoptosis of the infected cells (Sabbadini et al. 2012).

In our study, we analyzed the nucleotide and amino acid sequences of the genes encoding pili proteins which contribute to bacterial adherence to host cells, and also the gene encoding 67–72p protein involved in adhesion, colonization, and induction of the cell apoptosis in the early stage of infection, which should be used in a preliminary research for the finding of new vaccine antigens.

**Experimental**

**Materials and Methods**

**Bacterial strains.** In total, 10 *C. diphtheriae* non-toxigenic isolates were used in this study (Table I). Strains were isolated in Poland in 2010–2017 from patients with bacteremia, wound infection, septic arthritis, endocarditis, and serous cyst contents. The strain NCTC 13129 was used as the control strain.

| Strain | Biotype | Site of isolation | Year of isolation |
|--------|---------|-------------------|-------------------|
| 27/E   | mitis   | Serous cyst contents | 2010             |
| 40/E   | gravis  | Blood             | 2014             |
| 68/E   | gravis  | Endocarditis       | 2015             |
| 71/E   | gravis  | Wound             | 2015             |
| 73/E   | gravis  | Blood             | 2016             |
| 77/E   | gravis  | Wound             | 2016             |
| 78/E   | gravis  | Blood             | 2016             |
| 79/E   | gravis  | Blood and joint fluid | 2016         |
| 86/E   | gravis  | Blood             | 2017             |
| 89/E   | gravis  | Wound             | 2017             |
DNA isolation. *C. diphtheriae* strains were grown on Columbia agar with 5% sheep blood (BioMerieux) for 24 h at 37°C under aerobic conditions. Genomic DNA was isolated using a Wizard Genomic DNA Purification Kit (Promega) according to the Gram-positive bacteria procedure provided by the manufacturer.

Polymerase Chain Reaction (PCR). The oligonucleotide primers for amplification of the genes coding 67–72p protein and structural pil proteins (SpaC, SpaI, SapD) were designed based on the nucleotide sequence of the *C. diphtheriae* NCTC 13129 whole genome, available from GenBank under the number BX248353 (Table II).

The PCR reaction was conducted in a total volume of 25 μl and the reaction mixture contained 0.5 μl genomic DNA, 12.5 μl HotStarTaq Master Mix (Qiagen), and 1 μl of 10 μM solution of each primer (Table II). The cycling conditions were as follows: initial denaturation at 95°C/10 min and 29 cycles of denaturation at 94°C/1 min, primer annealing at 52°C/45 s, primer extension at 72°C/1 min and a final elongation at 72°C/10 min.

The PCR products were enzymatically cleaned using an Exo–BAP Mix kit (EURx), according to the manufacturer’s procedure and then sent for sequencing.

**Sequencing of fragment 4 of the gene encoding SpaC protein in strain 89/E.** Based on the result of the

| Gene | Primer | Sequence | Length of the amplified fragment |
|------|--------|----------|---------------------------------|
| 67–72p | 6772p1L | TGAATAATTATTTAAGGAGTTCCA | 695 bp |
| | 6772p1R | CAACCCACCACTAACCAGCAA | |
| | 6772p2L | CTGGTTTGCGTGCGTAGCA | 841 bp |
| | 6772p2R | AC gccACCTGTTGTTTGGC | |
| | 6772p3L | GAATCGTTGCAGCCCAAG | 699 bp |
| | 6772p3R | CCTTAAGACTGGGTGTTT | |
| | 6772p4L | CAACCCACCACTAACCAGCAA | 692 bp |
| | 6772p4R | TTCTGGTTGCTCCCTGGTCT | |
| | 6772p5L | TCAAGCCGGGATCCCAAGA | |
| | 6772p5R | TACGGTTGTCTGTTGAAAAGG | |
| spaI | SpaI1L | GCGGAAATCAACCCACCAAC | 600 bp |
| | SpaI1R | AACCGCTTACGATCCAAGGA | |
| | SpaI2L | ACACGGCCTTCCAACCTC | 482 bp |
| | SpaI2R | TGATATGAGGCGTGCTGCA | |
| sapD | SapD1L | TCGCGAAGGTAAGAAATACCTA | 698 bp |
| | SapD1R | CTTGTGATACCCACCGACCTT | |
| | SapD2L | GTCCAAAAACAGAGGCGGAAA | 814 bp |
| | SapD2R | GGGTTCAGTAAAAACCCAGTGG | |
| spaC | SpaC1L | GCCTACTCTCATGGGCAAGG | 824 bp |
| | SpaC1R | ACATGGCGATCTCCTGAAAT | |
| | SpaC2L | TCGTGAGGACGCTTAACAA | 838 bp |
| | SpaC2R | AACAGCAGCTTGGAGGCGAAA | |
| | SpaC3L | GGATCTCATAAGGGGGAATCG | 808 bp |
| | SpaC3R | TCAGCTGAGTTCTCTGTTCA | |
| | SpaC4L | CATTCTGGTTTGTCTCGTGA | 850 bp |
| | SpaC4R | GGTGTAGAAACGCGCTGCAAA | |
| | SpaC5L | CCAAAATCAACAGTTGGATATTACT | 850 bp |
| | SpaC5R | TTCTCTGTCATACAGCTGTCG | |
| | SpaC6L | CAAAATACCGATGGTCTGTTG | 845 bp |
| | SpaC6R | AGCTGATGAAATGGGAT | |
| | SpaC7L | CAAAAGGTTGCTTGGGCGATT | 687 bp |
| | SpaC7R | TCAGCCTGAAATGCTGCTT | |
| | SpaC8L | CTGGCATCTGGAATGTCTGAT | |
| | SpaC8R | ACCGAAACGGTGCTAGGGA | 578 bp |
first round of sequencing, new primers were designed to completely sequence a fragment of approximately 2000 bp, as follows: SpaC4L 5’-CATTCGTTTTTGTCCGTGA-3’ and SC1pR 5’-GAGCTCTTGAAGTTGCAGA-3’ (391 bp), SC2pL 5’-CCCAGAACCAGTTTGATGTAATG-3’ and SC2pR 5’-GGTATGCGGTCAGGTTTTT-3’ (687 bp), SC3pL 5’-CGCACAACATCGTGA-3’ and SC3pR 5’-ACAAGTGATTTCGCAGCA-3’ (850 bp), SC4pL 5’-ATGTTATCGCCGTATCCTG-3’ and SpaCAR 5’-GGTGTAGAAACGCCTCGA-3’ (764 bp). The PCR reaction was carried out according to the above conditions, the PCR products were then enzymatically purified and sent for sequencing.

Comparing the sequenced fragments and the translated nucleotide sequences for proteins. The sequences of the particular gene fragments were obtained as fluorograms. The BioEdit program was used for the gene assembly, alignment of sequences and comparative calculations. The translation of DNA into protein sequences was performed using the BLASTx program.

Analysis of 67–72p, SpaC, SpaI and SapD protein affinity to MHC classes I and II and linear epitopes. Amino acid sequences of *C. diphtheriae* NCTC 13129 reference strain available from the National Center for Biotechnology Information (NCBI) were used for analysis. To predict the cellular localization of the proteins, the software packages CELLO v.2.5 (Yu et al. 2006) and PSORTb version 3.0.2 (Yu et al. 2010) were used. The analysis for the transmembrane domains was performed using TMHMM Server v. 2.0. The MHC classes I and II binding prediction was performed using the immune epitope database (IEDB) (Kim et al. 2012) for the recommended human leukocyte antigen (HLA) allele set (Bui et al. 2005; Nielsen et al. 2003). Analysis has also been carried out using Propred-I (for MHC class I) and Propred (for MHC class II) regarding the number of alleles for which epitopes were found in the proteins analyzed (Singh and Raghava 2001; 2003). The prediction of linear B-cell epitopes has been carried out using the Bepipred Linear Epitope Prediction 2.0 (Jespersen et al. 2017). VaxiJen 2.0 was used for prediction of the protec- tive antigens. The results page on VaxiJen server creates lists of the selected target, the protein sequence, its prediction probability, and a statement of protective antigen or non-antigen, according to a predefined cut-off. Since more of the models had their highest accuracy at a threshold of 0.5, this threshold value was chosen for all bacterial models types (Doytchinova and Flower 2007).

**Results**

The isolates from invasive diseases and wound infections were included in the study because a wound can be a portal of entry for invasive infections. Moreover, future vaccines against non-toxigenic *C. diphtheriae* should protect against all kinds of infections. The analyses conducted in this study showed the variability of *C. diphtheriae* strains in terms of the nucleotide sequence of the genes encoding the 67–72p protein (99.37–100% average similarity) and structural proteins of pili SpaC (98.94–99.97% average similarity), SpaI (81.39–100% average similarity), and SapD (63.78–100% average similarity). The 27/E strain, the only representative of the mitis biotype, was the most different from the other strains tested. This strain did not have the gene coding for the Spa protein, while the average similarities of the nucleotide sequence of genes encoding the 67–72p, SapD, and SpaC proteins were 99.37%, 98.64%, and 38.94%, respectively when compared to the reference strain. Despite the significant variation, we managed to locate two fragments of the 67–72p protein (fragments No. 3 and 5), where the nucleotide sequences were 100% identical for all *C. diphtheriae* strains tested (the most conserved in the genome). In contrast, for *C. diphtheriae* biotype gravis we found as many as nine fragments located in the sequence of the gene encoding 67–72p protein (fragment No. 1, 3, 5) and the genes coding for the pili SapD (fragment No. 1), SpaI (fragment No. 2), and SpaC (fragment No. 2, 3, 7, 8), which were 100% identical in all nine strains tested (Table III).

The nucleotide sequences of the genes investigated were translated into the amino acid sequences. We revealed that the identified mutations resulted in a reading frame shift or were synonymous and nonsynonymous substitutions (Table IV).

In addition, we observed that a fragment of 1380 bp was inserted in the place of the gene encoding the SpaC protein in the 89/E strain (Table III). After sequencing the fragment No. 4 of spaC, we obtained the exact nucleotide sequence, which we compared to the available sequences in the GeneBank using the Nucleotide Basic Local Alignment Search Tool (BLASTn). In this way, we proved that the 1380 bp insertion sequence (IS) was transposed, which interrupted the continuity of the tested gene between 3072 bp and 3073 bp of the reference sequence of the NCTC 13129 strain. The result from the BLAST search revealed that the insertion fragment we detected was similar to the sequence of IS3 family transposase also identified in other *C. diphtheriae* strains but in different locations.

Analysis of the 67–72p, SpaC, SpaI and SapD protein affinity to MHC classes I and II, and linear B-cell epitopes in the first stage relied on the determination of the position of the proteins tested in the cell membrane and confirmation that all selected proteins were at least partially membranous or extracellular (Table V). Then, using the IEDB platform, it was observed that all proteins have high-affinity areas for MHC receptors of both classes and the fragments, which can be recog-
nized by antibodies. The output for the prediction of the high-affinity MHC binding peptides is typically given either in the units of a predicted affinity (IC$_{50}$ nanomolar) or as a percentile score reflecting the relative affinity of a selected peptide compared with a universe of random sequences. According to Paul et al. (2013), there are four categories of percentile ranks: 1) 0–0.30; 2) 0.30–1.25; 3) 1.25–5.0; and 4) 5.0–15.0. Their study proved that four pools of predicted peptides derived from the first two categories (0–0.30; 0.30–1.25) were immunogenic but finally, the transgenic mice in their study recognized only one peptide pool from the first category (0–0.30) (Paul et al. 2013). We can say that the smaller the percentile rank value, the higher the affinity. As for the IC$_{50}$ value, according to the IEDB Solutions Centre guidelines, IC$_{50}$ < 50 designs very high affinity, IC$_{50}$ < 500 – high affinity, and IC$_{50}$ < 5000 means low affinity (Fleri 2013). Accepting even the threshold of cutting off the percentile rank below 1 or IC$_{50}$ below 50, we still could have at least 100 to several hundred regions with high affinity for each of the proteins (Table VI, Table VII).

Table III
Comparison of the nucleotide sequences of all strains tested against the reference strain, given in percent (%).
The sequences are presented according to the analysed fragments.

| Target protein | Fragment | 27E | 40E | 68E | 71E | 73E | 77E | 78E | 79E | 86E | 89E |
|---------------|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 67–72p        | 1       | 98.35 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|               | 2       | 99.5  | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 99.87|
|               | 3       | 100   | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|               | 4       | 99.01 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 97.90|
|               | 5       | 100   | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| SapD          | 1       | 98.91 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|               | 2       | 98.37 | 27.52| 100 | 100 | 100 | 100 | 100 | 100 | 100 | 99.86|
| SpaI          | 1       | –     | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 62.77|
|               | 2       | –     | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| SpaC          | 1       | 72.78 | 99.87| 99.87| 99.87| 99.87| 99.87| 99.87| 99.87| 99.87| 99.87|
|               | 2       | 97.11 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|               | 3       | –     | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|               | 4       | –     | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 99.75|
|               | 5       | 91.47 | 99.21| 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|               | 6       | –     | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|               | 7       | 50.16 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|               | 8       | –     | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

* – In the fragment 4 of the gene encoding SpaC protein, the insertion sequence has been transposed

Table IV
Comparison of amino acid sequences of all strains tested against the reference strain, given in percent (%).

| Target protein | 27E | 40E | 68E | 71E | 73E | 77E | 78E | 79E | 86E | 89E |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 67–72p        | 99  | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| SapD          | 97  | 99  | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| SpaI          | –   | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 84  | 100 |
| SpaC          | 48  | 99  | 99  | 99  | 69  | 69  | 69  | 69  | 69  | 51  |

Table V
Extracellular regions of individual proteins.

| Region* | 67–72p | SpaC | SpaI | SapD |
|---------|--------|------|------|------|
| Start   | 44     | 138  | 235  | 310  |
| Stop    | 57     | 169  | 257  | 987  |

*Amino acid positions
and MHC-Ld) was omitted. The Proteasome Filter and Immunoproteasome filters were included in the analysis and for both, the threshold score of 4% was used. The ProPred1 cut-off threshold was also set at 4%. At the ProPred, 51 alleles related to MHC class II were considered. These were HLA-DR alleles. These molecules were encoded by DRB1 and DRB5 genes containing HLA DR1 (2 alleles), DR3 (7 alleles), DR4 (9 alleles),

| Target protein | Alleles       | Start | End | Peptide       | Percentile rank |
|----------------|---------------|-------|-----|---------------|-----------------|
| 67–72p         | HLA-A*02:06   | 6     | 14  | FTNDRFWSV     | 0.06            |
| 67–72p         | HLA-B*44:02   | 34    | 42  | SENDSSVEY     | 0.06            |
| 67–72p         | HLA-A*30:02   | 68    | 77  | RMASYWLDRY    | 0.06            |
| 67–72p         | HLA-B*44:02   | 9     | 17  | AEALSQVGI     | 0.07            |
| 67–72p         | HLA-A*02:06   | 27    | 36  | MILGALVPTV    | 0.07            |
| 67–72p         | HLA-A*68:01   | 1     | 9   | YAFTPALR      | 0.11            |
| 67–72p         | HLA-A*01:01   | 43    | 51  | DTDSSTYTY     | 0.11            |
| 67–72p         | HLA-A*01:01   | 48    | 56  | YTTTLSTLPY    | 0.11            |
| 67–72p         | HLA-B*44:03   | 34    | 42  | SENDSSVEY     | 0.11            |
| 67–72p         | HLA-B*57:01   | 58    | 66  | SSLAIGNAW     | 0.12            |
| SapD           | HLA-B*44:03   | 65    | 74  | AEWQELDTWW    | 0.06            |
| SapD           | HLA-B*07:02   | 4     | 13  | RIPIWAGIGAF   | 0.11            |
| SapD           | HLA-B*44:03   | 28    | 36  | KEGAYGLEY     | 0.11            |
| SapD           | HLA-A*68:01   | 47    | 55  | NVFKNNSN      | 0.12            |
| SapD           | HLA-B*40:01   | 10    | 18  | IEAQISGSL     | 0.17            |
| SapD           | HLA-A*24:02   | 70    | 79  | VWYAPQNIFF    | 0.18            |
| SapD           | HLA-A*68:01   | 24    | 32  | DTVGESAR      | 0.2             |
| SapD           | HLA-B*51:01   | 50    | 58  | YPLHISYLV     | 0.2             |
| SapD           | HLA-A*68:01   | 41    | 50  | EAPFQTVIPK    | 0.22            |
| SapD           | HLA-A*68:01   | 17    | 26  | EAYVKNGAFK    | 0.26            |
| SpaC           | HLA-A*11:01   | 658   | 666 | STNSVWIPK     | 0.06            |
| SpaC           | HLA-A*01:01   | 220   | 229 | LSDDKPFDLY    | 0.07            |
| SpaC           | HLA-B*53:01   | 233   | 241 | LPSEDYYYW     | 0.1             |
| SpaC           | HLA-A*68:02   | 191   | 199 | EVVELENNAV    | 0.1             |
| SpaC           | HLA-A*02:06   | 1859  | 1867| IVAAALWLV     | 0.11            |
| SpaC           | HLA-A*23:01   | 88    | 96  | PYRFGIYTF     | 0.11            |
| SpaC           | HLA-A*68:01   | 1436  | 1444| NTTSITYK      | 0.11            |
| SpaC           | HLA-A*31:01   | 363   | 371 | RFKNAARCQR    | 0.11            |
| SpaC           | HLA-B*44:02   | 1054  | 1063| AENTLSADAI    | 0.11            |
| SpaC           | HLA-A*23:01   | 1578  | 1587| SYTCTMPHLF    | 0.12            |
| SpaI           | HLA-A*30:01   | 2     | 11  | KKTHLFRIPA    | 0.08            |
| SpaI           | HLA-B*07:02   | 9     | 17  | IPAATTAAV     | 0.1             |
| SpaI           | HLA-B*07:02   | 147   | 155 | RPAEYRRTL     | 0.1             |
| SpaI           | HLA-B*57:01   | 109   | 117 | RSRLSDEVW     | 0.12            |
| SpaI           | HLA-A*30:02   | 129   | 137 | VTGLPMGVY     | 0.18            |
| SpaI           | HLA-A*02:01   | 137   | 145 | YLVSETPPA     | 0.2             |
| SpaI           | HLA-A*02:03   | 20    | 29  | LLASGPIASA    | 0.2             |
| SpaI           | HLA-A*02:06   | 153   | 161 | RTLDFLTV      | 0.21            |
| SpaI           | HLA-B*51:01   | 196   | 205 | FPPVESSYTL    | 0.24            |
| SpaI           | HLA-A*68:01   | 252   | 261 | LAIAGFLVQR    | 0.32            |

Table VI
MHC class I epitopes predicted from the target proteins.
Table VII
MHC class II epitopes predicted from the target proteins.

| Target protein | Alleles | Start | End | Peptide | Percentile rank |
|----------------|---------|-------|-----|---------|----------------|
| 67–72p HLA-DRB3*01:01 | 660 674 | DGSVDLYEFDENDPV | 0.01 |
| 67–72p HLA-DRB3*01:01 | 713 727 | MLARYHVDARDFFT | 0.01 |
| 67–72p HLA-DPA1*03:01/DPB1*04:02 | 15 29 | PQRRLTWLPILML | 0.01 |
| 67–72p HLA-DPA1*03:01/DPB1*04:02 | 173 187 | STFSVLLVVAFLIAL | 0.01 |
| 67–72p HLA-DRB1*07:01 | 49 63 | VDFRGVFNKVIATRI | 0.01 |
| 67–72p HLA-DPA1*01/DPB1*04:01 | 165 179 | LPA LRVVSTFSVLL | 0.01 |
| 67–72p HLA-DPA1*01/DPB1*04:01 | 267 281 | VISAVIAFSFSVIV | 0.01 |
| 67–72p HLA-DRB1*09:01 | 100 114 | PVVQYRAAVEKGYHR | 0.02 |
| 67–72p HLA-DRB3*01:01 | 712 726 | KMLARYHVDARDFF | 0.02 |
| 67–72p HLA-DPA1*01:03/DPB1*02:01 | 172 186 | VSTFSVLLVVAFLIA | 0.02 |
| SapD HLA-DRB3*01:01 | 181 195 | GKSIPSEHLKKNMYF | 0.01 |
| SapD HLA-DRB1*03:01 | 531 545 | PLHISYLVGDATIAR | 0.03 |
| SapD HLA-DQA1*04:01/DQB1*04:02 | 353 549 | SYLVGDATIARAKE | 0.09 |
| SapD HLA-DRB1*03:01 | 444 458 | PSDLALLPSKMTVSL | 0.12 |
| SapD HLA-DPA1*01/DPB1*04:02 | 616 630 | VQDEAVTTAAEWQEL | 0.13 |
| SapD HLA-DRB1*03:01 | 442 456 | DLPSDLALLPSKMTV | 0.13 |
| SapD HLA-DQA1*03:01/DQB1*03:02 | 615 629 | DVQDEAVTTAAEWQE | 0.16 |
| SapD HLA-DRB5*01:01 | 640 654 | LLGIIGAVGAVFLFR | 0.24 |
| SapD HLA-DQA1*04:01/DQB1*04:02 | 357 551 | LVGDATIARAKEILA | 0.27 |
| SpaC HLA-DRB3*01:01 | 1427 1441 | EHSVDPWLLNTTYSI | 0.01 |
| SpaC HLA-DRB3*01:01 | 1699 1713 | VVINNVTYTDATAE | 0.01 |
| SpaC HLA-DRB3*01:01 | 1804 1818 | EVTLNDYDADSGILT | 0.01 |
| SpaC HLA-DRB3*01:01 | 1450 1464 | IKDRSYSDWVQIAD | 0.02 |
| SpaC HLA-DRB1*03:01 | 783 799 | ADIVKYYVDNATKAR | 0.03 |
| SpaC HLA-DRB1*09:01 | 1845 1859 | NGYLRWLLAGAAG | 0.04 |
| SpaC HLA-DPA1*03:01/DPB1*04:02 | 21 35 | LAMMVSIIVLPLIA | 0.05 |
| SpaC HLA-DRB3*01:01 | 1432 1446 | PWLLNTTYSITYKCD | 0.07 |
| SpaI HLA-DRB3*01:01 | 147 161 | RPAEYRRTLDFLITV | 0.07 |
| SpaI HLA-DRB1*08:02 | 3 17 | KTLHRFRPATAAV | 0.18 |
| SpaI HLA-DRB3*01:01 | 151 165 | YRRTLDFLITVPAGM | 0.19 |
| SpaI HLA-DPA1*01/DPB1*04:01 | 248 262 | LGIALAAGFLVQR | 0.28 |
| SpaI HLA-DRB3*01:01 | 43 57 | ISDIRCDFGTSLTIL | 0.29 |
| SpaI HLA-DRB5*01:01 | 155 169 | LDIFLTVPGMRATAD | 0.42 |
| SpaI HLA-DRB1*09:01 | 95 109 | AGWDAAKALITQEAR | 0.44 |
| SpaI HLA-DRB1*11:01 | 50 64 | TGSILTIKRQPAFE | 0.44 |
| SpaI HLA-DQA1*01/02/DQB1*06:02 | 242 256 | VGLIAALGIALAAG | 0.48 |
| SpaI HLA-DPA1*02:01/DPB1*05:01 | 150 164 | EYRRTLDFLITVPAG | 0.6 |

DR7 (2 alleles), DR8 (6 alleles), DR11 (9 alleles), DR13 (11 alleles), DR15 (3 alleles), and DR51 (2 alleles). The cut-off threshold was set at 3%.

The target protein sequences were scanned for B-cell epitopes using the Bepipred Linear Epitope Prediction 2.0. The selected B-cell linear epitopes of the proteins analyzed are shown in Table IX.

The results obtained with the Vaxijen server also confirmed the possibility of using the proteins as antigens in vaccines (Table X).
Table VIII
Number of alleles for which epitopes were found in the proteins tested.

| Target protein | Number of MHC alleles of class I (per 40) | % Bound alleles MHC class I | Number of MHC alleles of class II (per 51) | % Bound alleles MHC class II |
|----------------|------------------------------------------|-----------------------------|---------------------------------------------|-----------------------------|
| 67–72p         | 40                                       | 100                         | 51                                          | 100                         |
| SpaC           | 37                                       | 92.5                        | 51                                          | 100                         |
| SpaI           | 34                                       | 85                          | 50                                          | 98                          |
| SapD           | 33                                       | 82.5                        | 51                                          | 100                         |

Table IX
B-cell epitopes predicted from the target proteins.

| Target protein | No. | Start | End | Peptide                           | Length |
|----------------|-----|-------|-----|-----------------------------------|--------|
| 6772p          | 1   | 4     | 20  | GFTRPAAPKRPOQRRLT                 | 17     |
| 6772p          | 2   | 48    | 53  | EVDFRDG                           | 6      |
| 6772p          | 3   | 87    | 111 | GRPDELEFFDPSVYQYRAAVEK           | 25     |
| 6772p          | 4   | 143   | 156 | NRQDFGVSDQQFGM                   | 14     |
| 6772p          | 5   | 194   | 211 | GGIRAGNQAAGVKGSITN               | 18     |
| 6772p          | 6   | 281   | 287 | VTKDLRI                           | 7      |
| 6772p          | 7   | 316   | 329 | SPNRAEKESEYISR                   | 14     |
| 6772p          | 8   | 337   | 369 | AYGITDDAVTYKDHWAGASEKVSATV5      | 33     |
| 6772p          | 9   | 381   | 410 | PTTFQQQQLNRFYFPKSLAMDRYVIDGEL    | 30     |
| 6772p          | 10  | 421   | 434 | DPNALKENQRDWIN                   | 14     |
| 6772p          | 11  | 452   | 467 | QVDEVARVGSARGGY                  | 16     |
| 6772p          | 12  | 474   | 490 | DLQTTTDEAQELGIVVY                | 17     |
| 6772p          | 13  | 498   | 507 | PVISATDGA                         | 10     |
| 6772p          | 14  | 514   | 541 | SENDSSVEYDTSTZYIQGKGGVNGN        | 28     |
| 6772p          | 15  | 562   | 566 | RNVGN                             | 5      |
| 6772p          | 16  | 573   | 581 | RDPREVPHN                         | 9      |
| 6772p          | 17  | 612   | 645 | TSLPYAERTXSLEATNQTTAQVGSQAQLVTDNV| 34     |
| 6772p          | 18  | 680   | 705 | GVFGTvKAISEELMNHLHYPED           | 26     |
| 6772p          | 19  | 714   | 749 | LARYHVDDARDFFTDNDRFWSVPSDPSTEGQKDVAQ| 36    |
| 6772p          | 20  | 760   | 763 | DTGK                              | 4      |
| 6772p          | 21  | 773   | 777 | RQLQR                             | 5      |
| 6772p          | 22  | 803   | 837 | TDTLTGQPKQAQDMVMSDDQIADNTLWKDTVNL | 35     |
| 6772p          | 23  | 861   | 868 | RKNQASAF                          | 8      |
| 6772p          | 24  | 896   | 944 | GIDPKQEADQDLGEAKEYLKTPESQRIDKPDKEKAPSTPSAPASGTTGE | 49 |
| 6772p          | 25  | 956   | 976 | LQSANKNGSNEYGRALDELDK            | 21     |
| SpaC           | 1   | 38    | 49  | ANAEPLPKKEFE                    | 12     |
| SpaC           | 2   | 64    | 69  | SLSASD                           | 6      |
| SpaC           | 3   | 100   | 113 | SPAAGNKNPFTPVSL                 | 14     |
| SpaC           | 4   | 131   | 146 | MPARIENKKSPNGGT                 | 16     |
| SpaC           | 5   | 176   | 184 | PTWDDNGRNR                      | 9      |
| SpaC           | 6   | 225   | 228 | PFDL                             | 4      |
| SpaC           | 7   | 231   | 234 | PILP                             | 4      |
| SpaC           | 8   | 246   | 254 | WKIDRSLTG                       | 9      |
| SpaC           | 9   | 324   | 332 | PSIETDKNG                      | 9      |
| SpaC           | 10  | 355   | 358 | TGDQ                             | 4      |
| SpaC           | 11  | 371   | 387 | RYSYQQAPTDPKMTSD                | 17     |
| SpaC           | 12  | 417   | 432 | KVNVTQLLEELNQ                | 16     |
| Target protein | No. | Start | End | Peptide | Length |
|----------------|-----|-------|-----|---------|--------|
| SpaC           | 13  | 455   | 468 | GVHNGESKEIGKVA | 14     |
| SpaC           | 14  | 478   | 507 | VTPKVDDSIRMKLTTWSSENTTADANQDNG | 30     |
| SpaC           | 15  | 512   | 522 | KADTDAFKNKK | 11     |
| SpaC           | 16  | 531   | 537 | NYEAQTA | 7      |
| SpaC           | 17  | 545   | 561 | JINRDKIPATKLPKFP | 17     |
| SpaC           | 18  | 569   | 591 | VPVPNARPEHGGLPETNPYFVDS | 23     |
| SpaC           | 19  | 601   | 610 | SIEIGPFPPYG | 10     |
| SpaC           | 20  | 619   | 659 | ARLPNVQADAKIPFGSLKTEWNSICFGNTIDNSQDCST | 41     |
| SpaC           | 21  | 664   | 672 | IPKPGQYSL | 9      |
| SpaC           | 22  | 676   | 684 | NTYTRELAS | 9      |
| SpaC           | 23  | 690   | 702 | TVSGDASDLNTSNH | 13     |
| SpaC           | 24  | 712   | 731 | DSGVEVYSDNIVVKKGSRQ | 20     |
| SpaC           | 25  | 746   | 754 | EKVPGQYSL | 9      |
| SpaC           | 26  | 761   | 769 | PFHLRSTAA | 9      |
| SpaC           | 27  | 779   | 786 | NTAKQRQA | 8      |
| SpaC           | 28  | 792   | 812 | KKVHKDKDTSpekISADALT | 21     |
| SpaC           | 29  | 819   | 846 | CTVPGVETPRKVLKTVSDNQTVEFGNFP | 28     |
| SpaC           | 30  | 857   | 861 | TEAPA | 5      |
| SpaC           | 31  | 881   | 885 | TPINK | 5      |
| SpaC           | 32  | 891   | 895 | FENAR | 5      |
| SpaC           | 33  | 904   | 948 | VLDGDMPQALVDQIPSSTTVNVACSTIGNHSITLQKDEQKAVPGV | 45     |
| SpaC           | 34  | 957   | 968 | SEETYTPITGATH | 12     |
| SpaC           | 35  | 971   | 991 | HWIKGELLEVADSTDTINPN | 21     |
| SpaC           | 36  | 1001  | 1007 | HYEDITAV | 7      |
| SpaC           | 37  | 1012  | 1037 | TKVRVQIDYQVNDVSNKNAVURP | 26     |
| SpaC           | 38  | 1043  | 1052 | RYRCEINGQV | 10     |
| SpaC           | 39  | 1059  | 1073 | SADAINATGATKVPRG | 15     |
| SpaC           | 40  | 1079  | 1131 | EEESSSVSLSNATLSHVEFFYHTGTKTNDKASVAINSDHNRLDATNTFTLTGTS | 53     |
| SpaC           | 41  | 1135  | 1146 | KKKVDGEGVSTI | 12     |
| SpaC           | 42  | 1157  | 1164 | RCTLGDWK | 8      |
| SpaC           | 43  | 1174  | 1188 | FDSAEHSVSDKIPVG | 15     |
| SpaC           | 44  | 1195  | 1204 | EDSEKAQEPN | 10     |
| SpaC           | 45  | 1210  | 1240 | RWHTHDSTNGWGDTEAACENHAACVEDPKNE | 31     |
| SpaC           | 46  | 1250  | 1255 | NKEFN | 6      |
| SpaC           | 47  | 1276  | 1288 | KVLTNDGPELAGK | 13     |
| SpaC           | 48  | 1298  | 1346 | TDPRFAGSALDKHISPDPITITVALNAKQGQRSASYQVADERHSVEVPV | 49     |
| SpaC           | 49  | 1357  | 1360 | IALY | 4      |
| SpaC           | 50  | 1378  | 1401 | AVQRTSSNSASARFYTEKQENNFG | 24     |
| SpaC           | 51  | 1409  | 1413 | DIYRP | 5      |
| SpaC           | 52  | 1424  | 1437 | AKPEHSVDPWLLNT | 14     |
| SpaC           | 53  | 1443  | 1483 | YKCDDPYIKDR5YSNDVDIQADAEKPTFADPTAHVKIPA | 41     |
| SpaC           | 54  | 1492  | 1498 | NTEGHLP | 7      |
| SpaC           | 55  | 1506  | 1555 | DETNKAFAEGHEKRYSTFPEIKDVLSEPEETFTNSYVMPRILS | 50     |
| SpaC           | 56  | 1560  | 1569 | VEGDPGHAVI | 10     |
| SpaC           | 57  | 1582  | 1605 | TMPHLPFPNPNSMQEVEGNKTVARG | 24     |
| SpaC           | 58  | 1614  | 1622 | TWRSPEVPI | 9      |
| SpaC           | 59  | 1630  | 1643 | EEDDPALRTKLENN | 14     |
| SpaC           | 60  | 1645  | 1687 | LRMVPTLVFPTERAGAASAPVPIPLTDRTPEYNTEPLQMPES | 43     |
Table IX. Continued.

| Target protein | No. | Start | End   | Peptide                      | Length |
|----------------|-----|-------|-------|------------------------------|--------|
| SpaC           | 61  | 1718  | 1723  | ADNSPL                       | 6      |
| SpaC           | 62  | 1734  | 1755  | GENGQRKELPEVADAPAKSAKS       | 22     |
| SpaC           | 63  | 1808  | 1825  | DNYDADSLITVEHPOQGK           | 18     |
| SpaC           | 64  | 1837  | 1842  | STLPLT                       | 6      |
| SapD           | 1   | 23    | 72    | PVSASEDAALDATGHHKGEPAFGVTIPKGTTYRSDGKEVPHPCVDRLKIG | 50 |
| SapD           | 2   | 86    | 96    | YSVKEPATDLPH                 | 11     |
| SapD           | 3   | 104   | 113   | DGQQVVPQES                   | 10     |
| SapD           | 4   | 122   | 145   | AGEDGEELSRIIPDDEESFLGK       | 11     |
| SapD           | 5   | 157   | 162   | IFANG                        | 6      |
| SapD           | 6   | 174   | 190   | DPHHEPKGKDSIPEHLD            | 17     |
| SapD           | 7   | 224   | 234   | SNDEELKTIEY                  | 11     |
| SapD           | 8   | 264   | 269   | AFKVKT                       | 6      |
| SapD           | 9   | 281   | 350   | DDEVGPLEGTTTNLNKTITPLKDLKDATNEPPTDPSEKKKPPRPEKGHSETSSPSA LDDSIERAWKLTGTPK | 70 |
| SapD           | 10  | 371   | 380   | TVINREGKKY                    | 10     |
| SapD           | 11  | 392   | 418   | SGGDQGPPLVKTDSWKDRIEAQISGL   | 27     |
| SapD           | 12  | 441   | 451   | EDLPSDALLPD                  | 11     |
| SapD           | 13  | 525   | 529   | GQKES                        | 5      |
| SapD           | 14  | 542   | 606   | TIARAKEILAGEKLGSLKKKPQKEKETTPASVQNKSKGHKNDTGVQGESARK RQQLAATSUGSDTN | 65 |
| SapD           | 15  | 624   | 632   | AAEQWQELDT                   | 9      |
| SpaI           | 1   | 22    | 50    | ASPIASADSRTITGTADGLNISDIRCDT | 29     |
| SpaI           | 2   | 55    | 75    | LIKRPPAAFEGVDKADLPAGT        | 21     |
| SpaI           | 3   | 86    | 124   | IEGIDLTKQAGWDAAKTIQEARSLSDEVKAVSIRD | 39 |
| SpaI           | 4   | 144   | 153   | PAKRPAYRR                    | 10     |
| SpaI           | 5   | 166   | 174   | RTADGNVAS                    | 9      |
| SpaI           | 6   | 186   | 242   | TDDLPLPTVPVPFPVESSVTILTPPSPVPGTPKPGPDLEKFRKEVTDRLGNT | 57 |
| SpaI           | 7   | 263   | 266   | KKNF                         | 4      |

Table X
Prediction of the protective antigens from the VaxiJen server.

| Protein | Overall Prediction for the Antigen |
|---------|-----------------------------------|
| 6772p   | 0.5123                            |
| SpaC    | 0.6757                            |
| SpaI    | 0.5504                            |
| SapD    | 0.5544                            |

Discussion

The huge success of vaccination against diphtheria almost enabled the elimination of the disease in Europe and other developed countries. However, in many countries with high vaccination coverage, i.e. France, Italy, Switzerland, Germany, and Canada an increase in non-toxigenic *C. diphtheriae* infections has been observed. For example, Poland is a country where the last case of diphtheria was recorded in the year 2000 and where the vaccination level of over 95% is achieved (Zasada et al. 2010). The first case of non-toxigenic infection with *C. diphtheriae* biotype *gravis* was recorded in Poland in 2004, where this bacterium induced sepsis and endocarditis in a patient (Zasada et al. 2005) and since then, practically every year, several cases of invasive *C. diphtheriae* infections have been diagnosed. In northern Germany, the number of non-toxigenic *C. diphtheriae* infections increased from five in 2013 to 23 in 2016, and 24 in only the first half of 2017 (Dangel et al. 2018). In England and Wales, a dramatic increase in infections was recorded since 1986, peaking at almost 300 cases in the year 2000 (Edwards et al. 2011). These examples indicate that the development of a new vaccine against non-toxigenic *C. diphtheriae* infection is of very important and necessary demand.

In vaccine development, the potential virulence factors exposed on the surface of a pathogen are considered
as suitable antigens for an effective acellular vaccine. It has been shown that pili of Gram-positive bacteria play a direct role in the pathogenesis. For example, studies on S. pneumoniae have proved that those strains, which have a pili island, adhere better to lung epithelial cells than do the strains that lack this island. In the invasive disease model, the piliated strain is more virulent and has a competitive advantage over the pili-negative strain after the mixed intranasal infection. Infection with the piliated strain induces a stronger inflammatory response and a higher level of the tumor necrosis factor in the bloodstream of mice, which may be due to the higher adhesion of the pilated bacteria to the cells involved in the innate immune response and their detection by host cell pattern-recognition receptors (Barocchi et al. 2006). The pili proteins are used as antigens in vaccines, for example in some acellular pertussis vaccines (Mosley et al. 2016).

The adhesion of Corynebacterium to host cells was observed for the first time for C. renale pili, which caused agglutination of trypsinized sheep red blood cells (Honda and Yanagawa 1974). It was not until more than thirty years later that Mandlik and colleagues identified adhesins, which were involved in adherence to pharyngeal host cells – the minor pilins SpaB and SpaC of C. diphtheriae (Mandlik et al. 2007). Subsequent studies showed that wild type C. diphtheriae cells bind to human lung epithelial, laryngeal, and pharyngeal cells, whereas mutants that lacked SrtA (i.e. they did not polymerize the SpaA-type pili) showed more than a 90% lower ability to adhere to human pharyngeal cells. Moreover, mutants that lacked only the major pili subunit, SpaA, showed a 10% reduction in adherence to these cells. In contrast, mutants that lacked either of the minor pilin subunits, SpaB or SpaC, showed a 70–75% reduction in adhesion. In addition, the latex beads coated only with SpaB or SpaC were sufficient to adhere to the host pharyngeal cells, while the SpaA-coated beads did not bind. SpaB and SpaC are anchored in the cell wall as monomers independent of the pilus structures. It is likely that the long pili mediate the initial attachment, while the monomeric pilins on the surface of the bacteria participate in the formation of an adhesion zone allowing the delivery of toxins and other virulence factors and may even play a significant role in host cell signaling (Rogers et al. 2011).

In addition to pili, C. diphtheriae produces the 67–72p protein located on its surface, which is involved in colonization, induction of apoptosis, and epithelial cell necrosis that were once attributed exclusively to the action of the diphtheria toxin (Sabbadini et al. 2012). This finding was also confirmed by Cerdeño-Tárraga et al. (2003) who sequenced the genome of the British clinical isolate (strain NCTC13129 biotype gravis – used in our study as the reference strain) and proved that the recent acquisition of pathogenicity factors went beyond the toxin itself and included the fimbrial proteins and adhesins. The 67–72p can act as an invasive and apoptotic protein for C. diphtheriae strains. The ability to penetrate, survive and induce apoptosis in epithelial cells may explain the endurance and dissemination of C. diphtheriae (Sabbadini et al. 2012). Proteins, which act as adhesins were also detected among other bacteria, e.g., the occurrence of the extracellular Eap protein was confirmed to be involved in colonization of eukaryotic cells by S. aureus strains (Haggar et al. 2003).

Our research is based on reverse vaccination. This method relies on the sequencing of pathogen genomes and determination in silico the most likely protective antigens prior to conducting experiments to prove this. Originally, this method was used to identify antigens as probable candidate vaccines against serogroup B meningococci (Christensen et al. 2013). In another study, Droppa-Almeida et al. (2018) used several available bioinformatics tools to design the efficient immunodominant epitopes for the development of the peptide vaccine against C. pseudotuberculosis for sheep and goats. Thanks to this research, it was possible to highlight the importance of bioinformatics software in the design of vaccines, especially in the identification of appropriate vaccine candidates, immune-informatics analysis and design of peptide vaccine from multi-epitopes (Droppa-Almeida et al. 2018). Bioinformatics tools present a lot of advantages, such as speed and low cost, so we used them at particular stages of the research.

First, we selected the gene encoding the 67–72p protein and the three pili genes spaC, sapD and spaI as the genes most frequently detected in various C. diphtheriae isolates as it was reported in our previous study (Zasada et al. 2012). In the present study, we found two fragments of 594 bp and 215 bp in the nucleotide sequence of the gene coding for the 67–72p surface protein, which were identical in all analyzed strains of C. diphtheriae, and in total, nine identical in C. diphtheriae biotype gravis strains sequences of the genes encoding 67–72p protein and structural proteins pili SpaI, SpaC and SapD (Table III). The sequence stability of these fragments represents a first step toward being the potential vaccine candidates. The analysis of amino acid sequences of these fragments confirmed that the proteins tested are located in the membrane or cell wall and have a large extracellular part (Table V).

An effective vaccine should induce a protective and long-lasting immune response. Therefore, we carried out analyses of the affinity of the tested proteins to MHC classes I and II and linear B-cell epitopes. MHC class I presents antigens to CD8+ T-cells and MHC class II presents antigens to CD4+ T-cells. The antigens, which are recognized by CD4+ and/or CD8+ T-cell receptors,
have the potential to stimulate a long-lasting and cytotoxic immune response. B-cell epitopes can induce both primary and secondary immunity. We showed that, in each of the proteins, areas with high affinity to MHC receptors can be distinguished (Table VI, Table VII, Table VIII) and we localized B-cell epitopes from target proteins (Table IX). In addition, the Vaxijen server was used that is a reliable and consistent tool for predicting protective antigens of bacterial, viral and cancer origin. The results obtained also confirmed that the proteins tested by us could be interesting to use as antigens in vaccines (Table X).

Our studies have shown that in the genome of the 89/E strain, the insertion element of 1380 bp was transposed and attached to fragment 4 of the gene encoding the SpaC pilin protein. The process of transposition of ISs can inactivate genes (Trost et al. 2012). Mandlik et al. (2007) confirmed the reduced C. diphtheriae adhesive activity as the result of mutations at the base pili protein SpaB and at the tip pili protein SpaC of the SpaA-type pili. Premature stop codons in the continuity of the genes encoding the proteins responsible for the adhesion of bacteria to host cells inactivate them and limit the colonization process. We did not investigate the adhesive activity of the strain 89/E and, therefore, we can only posit the influence of the insertion on the adhesive properties of the strain based on the data published by other researchers.

Due to the fact that in many European countries the number of infections with non-toxigenic C. diphtheriae strains has recently increased, a key aspect of our research was the understanding of virulence factors other than the diphtheria toxin and identification of new vaccine targets. An important problem is also the understanding of the colonization process and in particular the mechanism of adhesion and structure of the proteins, which participate in this process. Blocking the 67–72p surface protein or pilus structural proteins could effectively prevent the adhesion of C. diphtheriae bacteria to host tissues, colonization, and infection development. Due to the comparison of the nucleotide sequences of the C. diphtheriae strains identification of the most conserved sequences in the genome and determination of the variability between strains was achieved. The conserved sequences identified in 67–72p, SpaC, SapD and Spal in our study are identical for all C. diphtheriae strains tested and contain the epitopes for B-cells and T-cells and will be used in further research on the construction of a new vaccine. The main limitation of this study is the small number of isolates investigated. However, the results obtained here support further studies with a larger number of isolates from different countries. Moreover, the results of in silico analysis should be confirmed by in vivo studies on an animal model. The new vaccine will act to inactivate the antigens responsible for the host cell colonization by C. diphtheriae strains and inhibit the development of infection.

**ORCID**

Klaudia Brodzik 0000-0002-6875-2570
Maciej Polak 0000-0002-2288-5650
Aleksandra A. Zasada 0000-0003-2774-0941

**Acknowledgments**

The authors wish to thank the Laboratory of Bacteriological Diagnostics, Department of Bacteriology and Biocontamination Control NIPH-NIH for providing us with several C. diphtheriae isolates.

**Funding**

This study was supported by the National Institute of Public Health – National Institute of Hygiene internal-grant for young researchers (No. 1/EMMML/2017). Dominik Strapagiel and Jakub Lach were supported by the Polish Ministry of Science and Higher Education, grant DIR/WK/2017/01.

**Conflicts of Interest**

The authors do not report any financial or personal connections with other persons or organizations, which might negatively affect the contents of this publication and/or claim authorship rights to this publication.

**Literature**

Barocchi MA, Ries J, Zogaj X, Hemsley C, Albiger B, Kanth A, Dahlberg S, Fernebro J, Moschioni M, Massignani V, et al. A pneumococcal pili influences virulence and host inflammatory responses. Proc Natl Acad Sci USA. 2006 Feb 21;103(6):2857–2862.

Belko J, Wessely DL, Malley R. Endocarditis caused by Corynebacterium diphtheriae: case report and review of the literature. Pediatr Infect Dis J. 2000 Feb;19(2):159–163.

doi:10.1090/0000-6454-200002000-00015 Medline

Bui HH, Sidney J, Peters B, Sathiamurthy M, Sinichi A, Burton KA, Mothé BR, Chisari FV, Watkins DI, Sette A. Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. Immunogenetics. 2005 Jun;57(5):304–314. doi:10.1007/s00251-005-0798-y Medline

Cerdeño-Tárraga AM, Efstratiou A, Dover LG, Holden MT, Pallen M, Bentley SD, Besra GS, Churcher C, James KD, De Zoyza A, et al. The complete genome sequence and analysis of Corynebacterium diphtheriae NCTC13129. Nucleic Acids Res. 2003 Nov15; 31(22):6516–6523. doi:10.1093/nar/gkg874 Medline

Christensen H, Hickman M, Edmunds WJ, Trotter CL. Introducing vaccination against serogroup B meningococcal disease: an economic and mathematical modelling study of potential impact. Vaccine. 2013 May;31(23):2638–2646.

doi:10.1016/j.vaccine.2013.03.034 Medline

Colombo AV, Hirata R Jr, Souza CA, Monteiro-Leal LH, Previato JO, Formiga LCD, Andrade FAB, Mattos-Guaraldi AL. Corynebacterium diphtheriae surface proteins as adhesions to human erythrocytes. FEMS Microbiol Lett. 2001 Apr;197(2):235–239.

doi:10.1111/j.1574-6989.2001.tb01069.x Medline

Dangel A, Berger A, Konrad R, Bischoff H, Sing A. Geographically Diverse Clusters of Nontoxicogenic Corynebacterium diphtheriae Infection, Germany, 2016–2017. Emerg Infect Dis. 2018 Jul;24(7):1239–1245. doi:10.3201/eid2407.172026 Medline
New Corynebacterium diphtheriae vaccine antigens

Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007 Dec;8(1):4. doi:10.1186/1471-2105-8-4 Medline

Droppa-Almeida D, Franceschi E, Padilha FF. Immune-informatics analysis and design of peptide vaccine from multi-epitopes against Corynebacterium pseudotuberculosis. Bioinform Biol Insights. 2018 Jan;12:1–9. doi:10.1177/175922117855337 Medline

Edwards B, Hunt AS, Hoskisson PA. Recent cases of non-toxigenic Corynebacterium diphtheriae in Scotland: justification for continued surveillance. J Med Microbiol. 2011 Apr 01;60(4):561–562. doi:10.1099/jmm.0.025643-0 Medline

Fleri W. Selecting thresholds (cut-offs) for MHC class I and II binding predictions [Internet]. IEDB Solutions Center; 2014 [cited 2019 January 22]. Available from https://help.iedb.org/hc/en-us/articles/1140981181-Selecting-thresholds-cut-offs-for-MHC-class-I-and-II-bind-predictions

Fricchione MJ, Deyro HJ, Jensen CY, Hoffman JE, Singh K, Logan LK. Non-toxigenic penicillin and cephalosporin-resistant Corynebacterium diphtheriae endocarditis in a child: a case report and review of the literature. J Pediatric Infect Dis Soc. 2014 Sep 01;3(3):251–254. doi:10.1093/pids/pit022 Medline

Gaspar AH, Ton-That H. Assembly of distinct pilus structures on the surface of Corynebacterium diphtheriae. J Bacteriol. 2006 Feb 15;188(4):1526–1533. doi:10.1128/JB.188.4.1526-1533.2006 Medline

Haggar A, Hussain M, Lönnies H, Herrmann M, Norrby-Teglund A, Folk JI. Extracellular adherence protein from Staphylococcus aureus enhances internalization into eukaryotic cells. Infect Immun. 2003 May 01;71(5):2310–2317. doi:10.1128/IAI.71.5.2310-2317.2003 Medline

Hirata R Jr, Souza SMS, Rocha-de-Souza CM, Andrade AFB. Characterization of surface saccharides in two Corynebacterium diphtheriae strains. FEMS Microbiol Lett. 1999 Nov 15;179(1):119–125. doi:10.1111/j.1574-6968.1999.tb13369.x Medline

Mattos-Guaraldi AL, Formiga LC, Pereira GA. Cell surface components and adhesion on Corynebacterium diphtheriae. Microbes Infect. 2000 Oct;2(12):1507–1512. doi:10.1016/S1286-4579(00)01305-8 Medline

Mattos-Guaraldi AL, Formiga LCD, Andrade AFB. Cell surface hydrophobicity of sucrose fermenting and nonfermenting Corynebacterium diphtheriae strains evaluated by different methods. Curr Microbiol. 1999b Jan;38(1):37–42. doi:10.1007/BF0006769 Medline

Mattos-Guaraldi AL, Formiga LCD, Andrade AFB. Trans-Sialidase activity for sialic acid incorporation on Corynebacterium diphtheriae. FEMS Microbiol Lett. 1998 Nov 15;168(2):167–172. doi:10.1111/j.1574-6968.1998.tb13269.x Medline

Moreira LO, Andrade AFB, Vale MD, Souza SMS, Hirata R Jr, Asad LMOB, Asad NR, Monteiro-Leal LH, Previo JO, Mattos-Guaraldi AL. Effects of iron limitation on adherence and cell surface carbohydrates of Corynebacterium diphtheriae strains. Appl Environ Microbiol. 2003 Oct 01;69(10):5907–5913. doi:10.1128/AEM.69.10.5907-5913.2003 Medline

Mosley JF 2nd, Smith LJ, Parke CK, Brown JA, LaFrance JM, Clark PK. Quadrad: vaccination against diphtheria, tetanus, pertussis, and poliomyelitis in children. P T. 2016 Apr;41(4):238–253. Medline

Nielsen M, Lundegaard C, Worsøe P, Lomeland SJ, Hommerson ME, Jensen J, Lund Ø. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci. 2003 May;12(5):1007–1017. doi:10.1002/pai.203904 Medline

Paul S, Weiskopf D, Angelo MA, Sidney J, Peters B, Sette A. HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. J Immunol. 2013 Dec 15;191(12):5831–5839. doi:10.4049/jimmunol.1302101 Medline

Proft T, Baker EN. Pili in Gram-negative and positive bacteria – structure, assembly and their role in disease. Cell Mol Life Sci. 2009 Feb;66(4):613–635. doi:10.1007/s00018-008-8477-4 Medline

Readnor-Robinson ME, Ton-That H. Assembly and function of Corynebacterium diphtheriae pili. In: Burkovski A, editor. Corynebacterium diphtheriae and related toxigenic species. Dordrecht (Netherlands): Springer; 2014. p. 123–141. Rogers EA, Das A, Ton-That H. Adhesion by pathogenic corynebacteria. Adv Exp Med Biol. 2011;715:91–103. doi:10.1007/978-94-007-0940-9_6 Medline

Sabbadini PS, Assis MC, Trost E, Gomes DLR, Moreira LO, dos Santos CS, Pereira GA, Nagao PE, Azvedo VAC, Hirata Júnior R, et al. Corynebacterium diphtheriae 67-72p hemagglutinin, characterized as the protein DIP0733, contributes to invasion and induction of apoptosis in HEp-2 cells. Microb Pathog. 2012 Mar;52(3):165–176. doi:10.1016/j.micpath.2011.12.003 Medline

Sauer F, Mulvey MA, Schilling JD, Martinez J, Hultgren SJ. Bacterial pili: molecular mechanisms of pathogenesis. Curr Opin Microbiol. 2000 Feb 13(1):65–72. doi:10.1016/S1369-5278(99)00053-3 Medline

Singh H, Raghava GPS. ProPred1: Prediction of promiscuous MHC class-I binding sites. Bioinformatics. 2003 May 22;19(8):1009–1014. Medline

Singh H, Raghava GPS. ProPred: prediction of HLA-DR binding sites. Bioinformatics. 2001 Dec 01;17(12):1236–1237. doi:10.1093/bioinformatics/17.12.1236 Medline

Sorani M, Telford JL. Relevance of pili in pathogenic streptococci pathogenesis and vaccine development. Future Microbiol. 2010 May;5(5):735–747. doi:10.21272/fmb.10.37 Medline
Brodzik K. et al.

246

Swierczynski A, Ton-That H. Type III pilus of corynebacteria: pilus length is determined by the level of its major pilin subunit. J Bacteriol. 2006 Sep 01;188(17):6318–6325. doi:10.1128/JB.00606-06 Medline

Thanassi DG, Saulino ET, Hultgren SJ. The chaperone/usher pathway: a major terminal branch of the general secretory pathway. Curr Opin Microbiol. 1998 Apr;1(2):223–231. doi:10.1016/S1369-5274(98)80015-5 Medline

Ton-That H, Schneewind O. Assembly of pili in Gram-positive bacteria. Trends Microbiol. 2004 May;12(5):228–234. doi:10.1016/j.tim.2004.03.004 Medline

Ton-That H, Schneewind O. Assembly of pili on the surface of Corynebacterium diphtheriae. Mol Microbiol. 2003 Nov;50(4):1429–1438. doi:10.1046/j.1365-2958.2003.03782.x Medline

Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. Proteins. 2006 Aug 15;64(3):643–651. doi:10.1002/prot.21018 Medline

Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinap SC, Ester M, Foster IJ, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010 Jul 1;26(13):1608–1615. doi:10.1093/bioinformatics/btq249 Medline

Zasada A, Zaleska M, Podlasin R, Seferynska I. The first case of septicemia due to nontoxigenic Corynebacterium diphtheriae in Poland: case report. Ann Clin Microbiol Antimicrob. 2005;4(1):8. doi:10.1186/1476-0711-4-8 Medline

Zasada AA, Baczewska-Rej M, Wardak S. An increase in nontoxigenic Corynebacterium diphtheriae infections in Poland – molecular epidemiology and antimicrobial susceptibility of strains isolated from past outbreaks and those currently circulating in Poland. Int J Infect Dis. 2010 Oct;14(10):e907–e912. doi:10.1016/j.ijid.2010.05.013 Medline

Zasada AA, Formińska K, Rzeczkowska M. [Occurrence of pili genes in Corynebacterium diphtheriae strains] (in Polish). Med Dosw Mikrobiol. 2012;64(1):19–27. Medline

Zasada AA. Nontoxigenic highly pathogenic clone of Corynebacterium diphtheriae, Poland, 2004–2012. Emerg Infect Dis. 2013 Nov;19(11):1870–1872. doi:10.3201/eid1911.130297 Medline