scientific reports

Check for updates

OPEN Identification of linear epitopes on the flagellar proteins of *Clostridioides difficile*

A. Razim[1✉], K. Pacyga[1], P. Naporowski[2], D. Martynowski[3], A. Szuba[4,5], A. Gamian[2] & S. Górska[1]

Clostridioides difficile (*C. difficile*) is an opportunistic anaerobic bacterium that causes severe diseases of the digestive tract of humans and animals. One of the possible methods of preventing *C. difficile* infection is to develop a vaccine. The most promising candidates for vaccine antigens are the proteins involved in the adhesion phenomena. Among them, the FliC and FliD are considered to be suitable candidates. In this paper, the FliC and FliD protein polypeptide epitopes were mapped in silico and by using PEPSCAN procedure. We identified four promising epitopes: [117]QRMRTLS[123], [205]MSKAG[209] of FliC and [226]NKVAS[230], [306]TTKKPKD[312] of FliD protein. We showed that [117]QRMRTLS[123] sequence is not only located in TLR5-binding and activating region, as previously shown, but forms an epitope recognized by *C. difficile*-infected patients' antibodies. [205]MSKAG[209] is a *C. difficile*-unique, immunogenic sequence that forms an exposed epitope on the polymerized flagella structure which makes it a suitable vaccine antigen. [226]NKVAS[230] and [306]TTKKPKD[312] are well exposed and possess potential protective properties according to VaxiJen analysis. Our results open the possibility to use these epitopes as suitable anti-*C. difficile* vaccine antigens.

C. difficile is an opportunistic anaerobic bacterium that causes severe diseases of the digestive tract of humans and animals. It is a Gram-positive, rod-shaped pathogen producing dangerous to health toxins (TcdA and TcdB) and spores that can survive in high temperatures and generally used cleaning agents[1]. *C. difficile* is responsible for colitis and the occurrence of acute life-threatening diarrhea especially in hospitalized elderly and patients undergoing antibiotic treatment. Since the number of hypervirulent *C. difficile* strains as well as the strain resistance towards commonly used antibiotics is increasing, this bacteria is a serious problem that requires development of new prevention and treatment methods[2,3].

One of the possible methods of preventing *C. difficile* infection (CDI) is to induce by vaccination specific antibodies directed against molecules involved in the adhesion process[4]. Blocking the binding sites on the surface of bacteria prevents the adhesion of pathogens to the epithelial cells of host and stops the development of infection. It has been shown that protein components of the surface of *C. difficile* have immunomodulatory properties[5–7]. Moreover, they contain conserved regions in their structure, thus they are ideal candidates for vaccine components/antigens. The most promising results of the immunization studies and serological analysis were obtained for proteins: Cwp66, Cwp84, FliC, FliD and Fbp68[5,8,9]. Simultaneously these antigens can be used to obtain protective and therapeutic antibodies.

Flagellum was shown to have an important role in adhesion of various pathogens and is one of the vital virulence factors[10–12]. Flagellin, the most abundant protein in the flagellum, is stimulating host immune system by interaction with Toll-Like-Receptor 5 (TLR5) and is considered to be a suitable vaccine adjuvant[13]. In *C. difficile*, the flagellum is composed of FliC (flagellin) and FliD (cap) proteins and both of them bind murine mucus[14] and are implicated in the process of biofilm formation[15]. In contrast to non-flagellated or non-toxigenic strains of *C. difficile*, the flagellated strains induce inflammatory response synergistically with toxins in CDI mouse models[16]. It is well known that *C. difficile*-infected patients develop a robust antibody response to FliC and FliD proteins. What is particularly interesting is that the level of antibodies against flagella is lower in the *C. difficile*-infected patient group than in the control group which suggests their possible protective role[8]. It was previously shown

[1]Laboratory of Immunobiology of Microbiome, Hirszfeld Institute of Immunology and Experimental Therapy, PAS, Wroclaw, Poland. [2]Laboratory of Medical Microbiology, Hirszfeld Institute of Immunology and Experimental Therapy, PAS, Wroclaw, Poland. [3]Laboratory of Genomics and Bioinformatics, Hirszfeld Institute of Immunology and Experimental Therapy, PAS, Wroclaw, Poland. [4]Division of Angiology, Wroclaw Medical University, Wroclaw, Poland. [5]Department of Internal Medicine, 4th Military Hospital in Wroclaw, Wroclaw, Poland. ✉email: agnieszka.razim@hirszfeld.pl

that recombinant FliC is immunogenic and protective in a murine model of CDI and partially protective in CDI hamster model[17]. The study of restriction fragment length polymorphism of 17 clinical isolates of *C. difficile* showed high similarity of protein sequences of FliD and FliC between isolates[9]. Both flagella-building proteins of *C. difficile* are good candidates for use in vaccines in combination with suitable adjuvants due to their conservative nature and the ability of inducing a strong immune response.

To ensure vaccine safety each antigen has to be thoroughly characterized in terms of its cross-reactivity and possible autoimmunizing properties[7]. Using epitope-based vaccines is a well-recognized approach that allows to obtain safe and effective formulations. FliC and FliD have been proposed to be suitable as anti-*C. difficile* antigens. So far, there was no study designed to explore their epitopes using empirical methods. In this paper we describe the process of epitope mapping of two flagellar proteins from *C. difficile* using three types of patient sera. We identified the shortest amino acid sequences recognized by patient sera, tested their cross-reactivity with other clinically-relevant flagellated bacterial strains. We also bioinformatically defined the localization of identified epitopes in the assembled flagellum and evaluated the possibility of using flagellar epitopes as vaccine antigens.

Methods

Blood sera.
Human blood sera. Serum samples from *C. difficile*-infected patients were provided by 4th Military Clinical Hospital in Wrocław. Umbilical cord blood samples ($n = 10$) and peripheral blood samples from healthy volunteers ($n = 16$) were provided by internal collection of Hirszfeld Institute of Immunology and Experimental Therapy. All sera samples were collected and used under the written approval of Bioethics Committee of the Medical University of Wroclaw (no. KB-631/2015). An informed consent was obtained from all of the participants. Experiments were conducted in accordance with the Helsinki Declaration, 1975. In total, serum samples from sixteen patients suffering first episode of CDI were collected. Patients were diagnosed based on the following symptoms: three or more loose stools within 24 h, fever, abdominal pain, and positive test results for the presence of glutamate dehydrogenase, as well as CD toxins (C. Diff Quik Chek Complete; TECHLAB, Inc., Blacksburg, VA, USA)[18]. Within groups, the sera were pooled, aliquoted, and stored at −20 °C until further used.

Animal blood sera. Serum samples of rabbits immunized with whole acetone-inactivated bacteria: *Escherichia coli* O56 PCM 2372 (*E. coli*), *Citrobacter freundii* PCM 1506 (*C. freundii*), *Hafnia alvei* PCM 1203 (*H. alvei*), *Shigella sonnei* (*S. sonnei*) were used in this manuscript. All of the above sera were already used in other experiments[19,20] and were obtained from the sera collection of Hirszfeld Institute of Immunology and Experimental Therapy.

Multiple sequence alignment.
All used sequences were obtained from National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) protein database[21]. Multiple sequence alignment and analysis of sequence homology were performed using Clustal Omega program (https://www.ebi.ac.uk/Tools/msa/clustalo/)[22] and BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Prediction of linear antigenic epitopes.
The FliC (NCBI accession number: CBE01876) and FliD (NCBI accession number: CBE01872) sequences of *C. difficile* R20291 strain used for epitope prediction were obtained from NCBI protein sequence database. For prediction of pentadeca- and hexadecapeptide linear antigenic epitopes EPMLR tool and SVMTrip were used[23,24].

Peptide synthesis.
NCP Block of 96 hydroxypropylmethacrylate pins and F-moc protected amino acids were obtained from Mimotopes (Clayton, Victoria, Australia). Chemicals used for synthesis and side chain deprotection like piperidine, dimethylformamide (DMF), trifluoracetic acid, diisopropylcarbodiimide (DIC) were of analytical grade, purchased from Merck Millipore (Burlington, MA, USA). Reagents like 2-mercaptoethanol, ethanedithiol, anisole, 1-hydroxybenzotriazole (HoAt), N,N-diisopropylethylamine (DIEA) and bromophenol blue were purchased from Sigma-Aldrich (Saint Louis, MO, USA).

A total of 20 FliC 15-amino acid- and 32 FliD 16-amino acid-long peptides covering predicted epitopes were synthesized. Peptides were synthesized according to standard protocol[25] with slight changes[6,26].

Pin-bound modified ELISA.
In order to test the interaction between serum antibodies and pin-bound synthetic peptides a modified ELISA was performed. 96-well polystyrene plates were used for all ELISAs. First, pins were incubated for 1 h in a solution of 1% bovine serum albumin (BSA, Sigma Aldrich) in TBS-T (Tris-buffered saline with 0.05% Tween20) at room temperature. Patients' sera (or rabbit sera) in 1:1 000 dilution were used as primary antibodies source, incubated for 2 h at room temperature. In the next step pins were washed with TBS-T. Pins were incubated with secondary antibodies conjugated with alkaline phosphatase (anti-human 1:10,000 or anti-rabbit 1:30 000 IgG antibodies) for 1 h at room temperature (Sigma-Aldrich, cat. no A1543 and A3687). Pins were washed again and the color reaction was developed with AP Yellow for 30 min at room temperature. The reaction was stopped by removing pins from the solution and the colorimetric reaction was measured at 405 nm with a microplate reader (PowerWave HT, BioTek Instruments, Winooski, VT, USA). In the last step pins were regenerated by sonication for 10 min in disruption buffer preheated to 55 °C using ultrasonic bath (Cole Parmer, Bunker Court Vernon Hills, IL, USA). Pins were washed with MilliQ water preheated to 55 °C for 5 min. Pins intended for storing were additionally washed with methanol for 2 min and air-dry. All tests were performed at least in triplicate. The baseline for ELISA result analysis was counted as a mean of all readings obtained for *C. difficile*-infected patient group[26].

Recognition of minimal epitopes and essential amino acids. Peptide sequences which showed the highest antibody binding were mapped in detail to determine the shortest amino acid sequence needed for the activity. For each pentadeca- and hexadecapeptides a library of truncated sequences was synthesized. The library was prepared by synthesizing peptides with removed flanking residues of the original peptide.

In the next step, essential amino acids were recognized by alanine, glycine or valine scanning, the so-called alanine-walk[27]. Scanning was performed by synthesizing a set of peptides in which each of the following amino acids is replaced by an alanine, glycine or valine.

Immunogenicity and autoimmunoreactivity. Epitopes identified by human sera were further evaluated using VaxiJen v.2.0 which is an alignment-independent tool predicting protein antigenicity basing on its physical properties[28]. Since two epitopes were to short for VaxiJen analysis, we added glycine at the C-end of the peptide for this analysis. Moreover, we searched the autoepitope database in order to confirm that the epitopes of FliC and FliD are not related to any known auto-epitopes[29]. We also looked for similar sequences in the known human protein amino acid sequences using BLAST.

Epitope localization in flagella. A model of *C. difficile* FliC monomer was generated by Phyre2 server[30] with 100% confidence. A FliD monomer was obtained with SWISS-MODEL[31]. The models of FliC and FliD were superimposed on a cryo-EM structure of *Pseudomonas aeruginosa* flagellar filaments (pdb 5wk6) and on a flagellar cap hexamer from *Escherichia coli* (pdb 5h5v) respectively. The interactions between TLR5 and *C. difficile* FliC were generated by sequence alignment of *C. difficile* model and *Salmonella enterica* protein (pdb 3v47) using Pymol[32]. All superpositions and protein figures were prepared with Pymol. We characterized the localization of epitopes in these proteins. Analysis of the protein domains was performed with Predict Protein software[33].

Statistical analysis. All graphs were prepared and data analyzed with GraphPad Prism 9. Unless otherwise stated data were analyzed using two-way ANOVA ($p < 0.05$). All measurements were performed at least in a triplicate. Pearson correlation coefficient ($p < 0.05$) was used to analyze the link between sequence similarity of *C. difficile* FliC and FliD epitopes to flagellar sequences of other bacterial strains and their immunoreactivity.

Results

Mapping of flagellar amino acid sequences with patients' sera revealed highly immunoreactive regions.
We designed a set of peptides for chemical synthesis that covered whole proteins amino acid sequences and overlaid in silico predicted epitopes (Supplementary Tab. S1). We synthesized 20 peptides based on FliC sequence from *C. difficile* R20291 strain that were 15-amino acid long (Supplementary Tab. S2). We synthesized 32 peptides based on FliD sequence from *C. difficile* R20291 strain that were 16-amino acid long (Supplementary Tab. S3). Peptides were analyzed in terms of their immunoreactivity with three types of sera: *C. difficile*-infected patients, healthy controls and umbilical cord blood sera. All of the used sera were diluted 1:1 000 as described in M&M section.

Epitope mapping of pentadecapeptides synthesized based on FliC amino acid sequence resulted in two peptide candidates for detailed mapping which are NTSSIMSKAGITSST (FliC7) and NILQRMRTLSVQSSN (FliC17) (Fig. 1A). Overall, peptides were recognized at the highest level by IgG antibodies from sera obtained from patients during the initial symptomatic infection (*C. difficile*-infected patients). The same peptides were recognized at a much lower level by the sera of other groups (umbilical cord blood and healthy blood donors group). The baseline was calculated as the mean of all measurements obtained for *C. difficile*-infected patient group (mean absorbance = 1.41). Six of the tested peptides (FliC1, 2, 7, 10, 13, 17) were significantly more immunoreactive than the rest but not all of them were suitable for further analysis as vaccine epitopes because of their cross-reactivity with serum from rabbits immunized with other flagellated bacterial strains or high immunoreactivity with sera from the relapsed patient group (Fig. 2 and Supplementary Fig. S1) or because of their high immunoreactivity with sera from people without the history of CDI can be a result of some cross-reactivity. An interesting observation is that FliC11 which contains one of the bioinformatically predicted epitope in its sequence has a very low immunoreactivity. FliC7 and FliC17 were qualified by us for further detailed analysis.

Epitope mapping of hexadecapeptides synthesized based on FliD amino acid sequence resulted in two candidates for detailed mapping which were ASGNKVASVYGKNLEA (FliD17) and TTKKPKDYPPLTDAQK (FliD21) (Fig. 1B). As in the case of FliC, peptides were recognized at the highest level by *C. difficile*-infected patients' sera group, then umbilical cord blood sera and at least by the healthy controls. The calculated baseline for *C. difficile*-infected patient group was 1.34. Ten of synthesized peptides were significantly more immunoreactive than the rest and were analyzed in terms of their cross-reactivity with serum from rabbits immunized with other flagellated bacterial strains. FliD1, 2 and 3 which were designed to cover bioinformatically predicted epitopes had a very low immunoreactivity with all three groups of tested sera. FliD17 and FliD21 were qualified by us for further detailed analysis because of their high immunoreactivity with CDI patient sera and low immunoreactivity with healthy sera.

The level of amino acid sequence homology only partially determines peptide cross-reactivity.
Simultaneously with epitope mapping of FliC7, FliC17, FliD17 and FliD21 peptides, we analyzed the cross-immunoreactivity of synthesized *C. difficile* FliC and FliD peptides with sera from rabbits immunized with other flagellated bacteria strains. We used sera from rabbits immunized with *C. freundii*, *H. alvei*, *E. coli* and *S. sonnei* in order to decipher the potential impact of peptide homology in their immune-reactivity. For both proteins we selected a set of peptides with varying sequence homology to FliC (Supplementary Fig. S2) and FliD
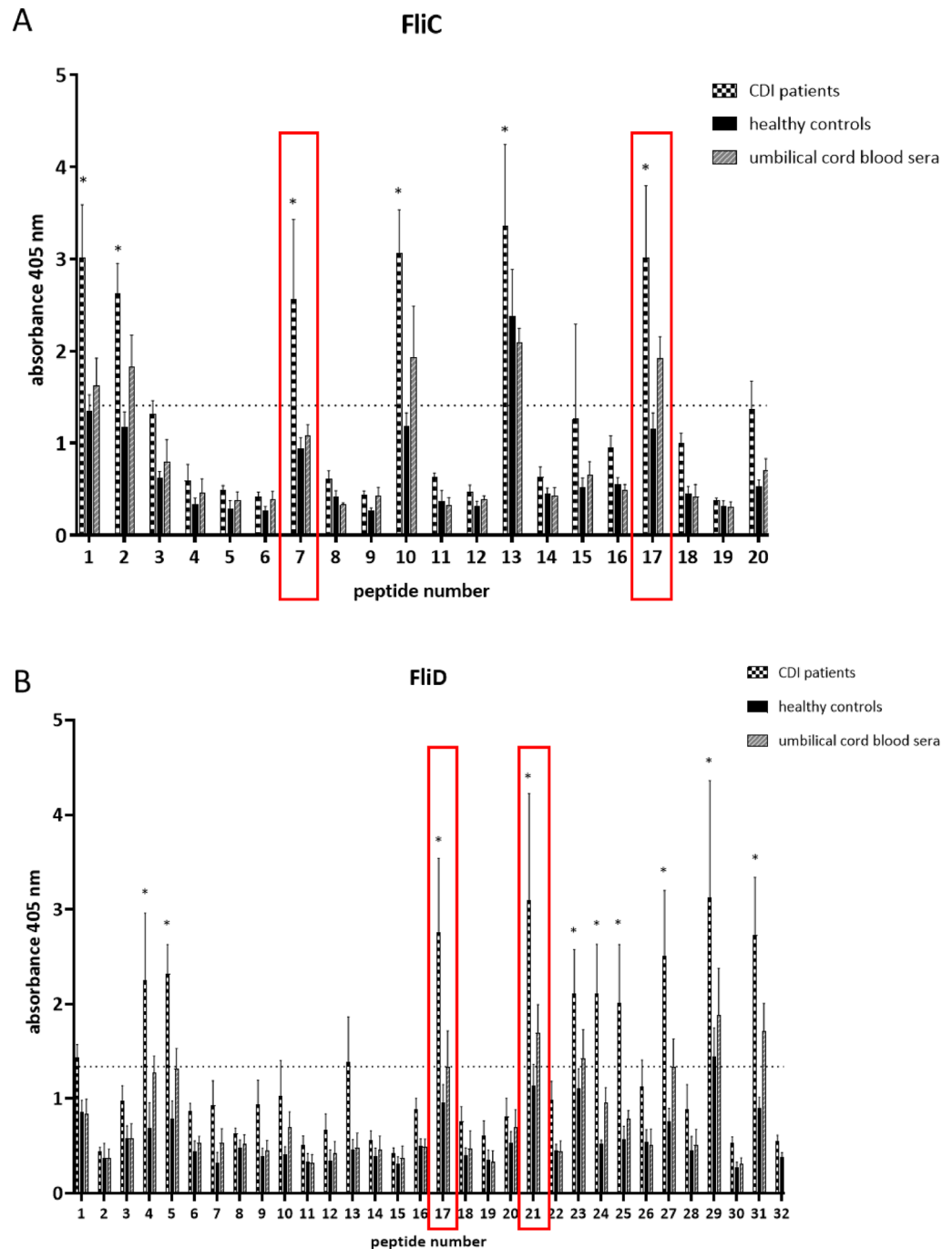
Figure 1. Immunoreactivity of peptides synthesized for mapping FliC (**A**) and FliD (**B**). Immunoreactivity was analyzed by pin-bound ELISA in which three groups of pooled sera: *C. difficile*-infected patients, healthy controls and umbilical cord blood sera were used in 1:1000 dilution. Anti-human IgG antibody conjugated to AP was used in 1:10,000 dilution. Assay was performed four times, the data shown is mean with ± SD. Statistical significance of the results obtained for *C. difficile*-infected patient group was calculated using two-way ANOVA in reference to the calculated baseline (1.41 for FliC and 1.34 for FliD) *$p < 0.0332$. Red brackets indicate peptides selected for further analysis.

(Supplementary Fig. S3) of the above bacterial strains and varying immunoreactivity. Cross-reactivity analysis performed by us shows that the level of amino acid sequence similarity is not always the best prognostic marker for possible cross-reactivity. For peptides based on *C. difficile* FliC we got a linear relationship between sequence similarity and cross-reactivity with *C. freundii* and *S. sonnei* vaccinated rabbits (Fig. 2a,d). However, when we counted the Pearson correlation coefficient for this data the visible relationship turned out to be statistically insignificant (Supplementary Tab. S4). Peptide FliC7 that had the lowest homology with flagellin of *C. freundii*
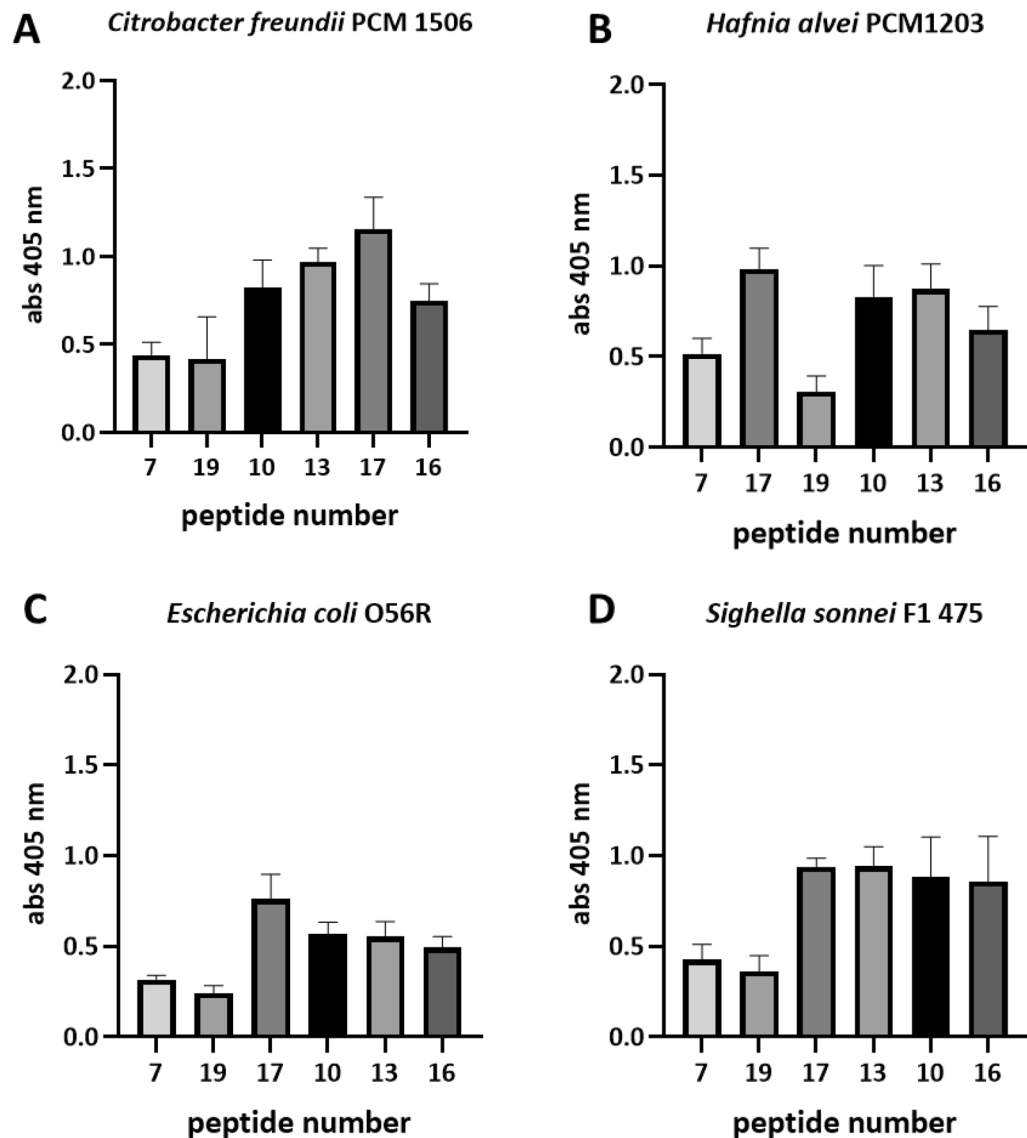
Figure 2. Cross-reactivity analysis of *C. difficile* FliC peptides and sera from rabbits immunized with other flagellated bacterial strains. *C. difficile* FliC peptides are organized from the ones with the lowest sequence homology to the flagellar sequences of bacteria used for rabbit immunizations to the ones with the highest sequence homology. Pin-bound ELISA was performed using rabbit sera (1: 1000) from animals immunized with whole acetone-inactivated bacteria. ELISA was performed at least three times. Data shown are means with ±SD.

(1/15 similar amino acids), *H. alvei* (1/15), *E. coli* (1/15) and *S. sonnei* (4/15) showed low cross-reactivity but most homological peptide FliC16 (9–10/15) was never the one with the highest level of cross-reactivity.

In case of *C. difficile* FliD peptides there is a visible and statistically significant relationship between sequence similarity and cross-reactivity of two tested sera—*E. coli* ($p = 0.0391$) and *S. sonnei* ($p = 0.0016$) (Supplementary Tab. S4). It needs to be mentioned that the overall homology between *C. difficile* FliD protein and the corresponding proteins from other bacterial strains is very low as well as the cross-reactivity for most of the tested sequences. FliD16 has no homological sequence in *C. freundii* and *H. alvei*; FliD 30 has no homological sequence in *E. coli*. The highest homology was determined for FliD20 (6/16 of similar amino acids for *E. coli* and *H. alvei*; 7/16 for *S. sonnei*; 9/16 for *C. freundii*) (Fig. 3).

Recognition of flagellar epitopes. We mapped epitopes in two peptides of each of the flagellar proteins FliC7 and FliC17 from *C. difficile* flagellin and FliD17 and FliD21 from *C. difficile* cap protein. We synthesized a library of truncated peptides for each of the above sequences and analyzed their immunoreactivity with two types of patients' sera (initial episode of CDI and healthy control sera). By this method, we were able to find the epitope in FliC7 which is [205]MSKAG[209] and in FliC17 which is [117]QRMRTLS[123] (Fig. 4A). In the next step we characterized the amino acids essential for antibody binding by "alanine walk" method. In epitope [205]MSKAG[209] each amino acid was substituted with valine since in the sequence there are both alanine and glycine, which
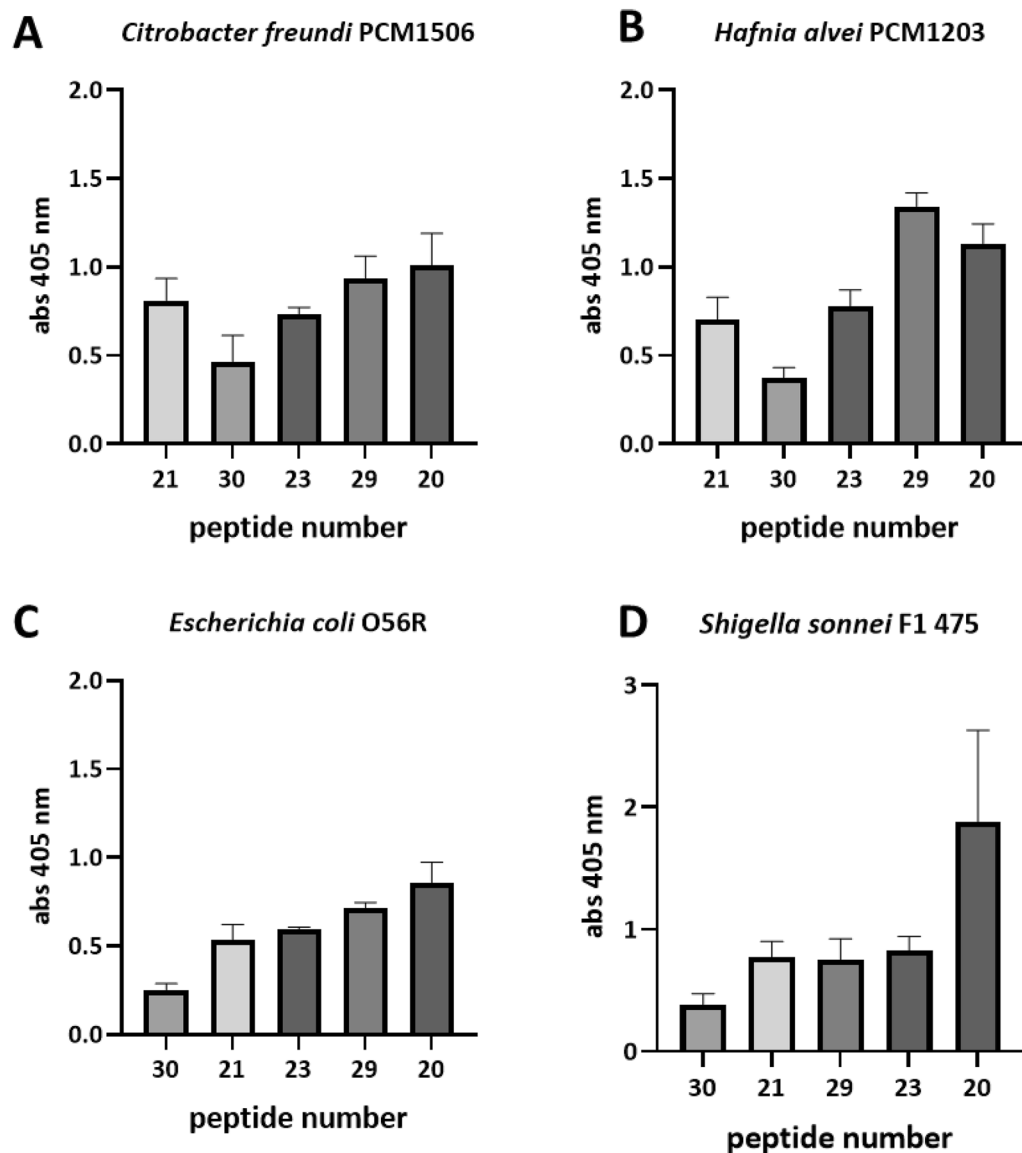
Figure 3. Cross-reactivity analysis of *C. difficile* FliD peptides with sera from rabbits immunized with other flagellated bacterial strains. *C. difficile* FliD peptides are organized from the ones with the lowest sequence homology to the flagellar sequences of bacteria used for rabbit immunizations to the ones with the highest sequence homology. Pin-bound ELISA was performed using rabbit sera (1: 1000) from animals immunized with whole acetone-inactivated bacteria. ELISA was performed at least three times. Data shown are means with ± SD.

are routinely used in this method. There is one essential amino acid in the [205]MSKAG[209] epitope which is lysine (Fig. 4B). Substitution of lysine in the [205]MS**K**AG[209] sequence results in total lost of epitope immunoreactivity. However, all of the other amino acids in the sequence are important for epitope reactivity since a substitution of each of them results in slight decrease in the level of immunoreactivity, especially in the case of serine. In epitope [117]QRMRTLS[123] there is no essential amino acid which substitution would lead to a total loss of epitope activity (Fig. 4B). However, there is a significant change in immunoreactivity when each of the two arginines in the sequences are substituted with alanine. Arginine as well as lysine belong to the group of positively charged amino acids.

By using the library of truncated peptides we have found epitopes in FliD17 and FliD21 which are [226]NKVAS[230] and [306]TTKKPKD[312], respectively (Fig. 5A). Alanine walk method showed that in [226]NKVAS[230] epitope there is one essential amino acid which is lysine (Fig. 5B). The substitution of the rest of the amino acids did not change significantly epitope immunoreactivity. However, these amino acids might be important for epitope recognition by antibodies. In [306]TT**KK**PKD[312], again lysins are the most important for antibody binding. A substitution of each of them with alanine significantly changed epitope immunoreactivity.
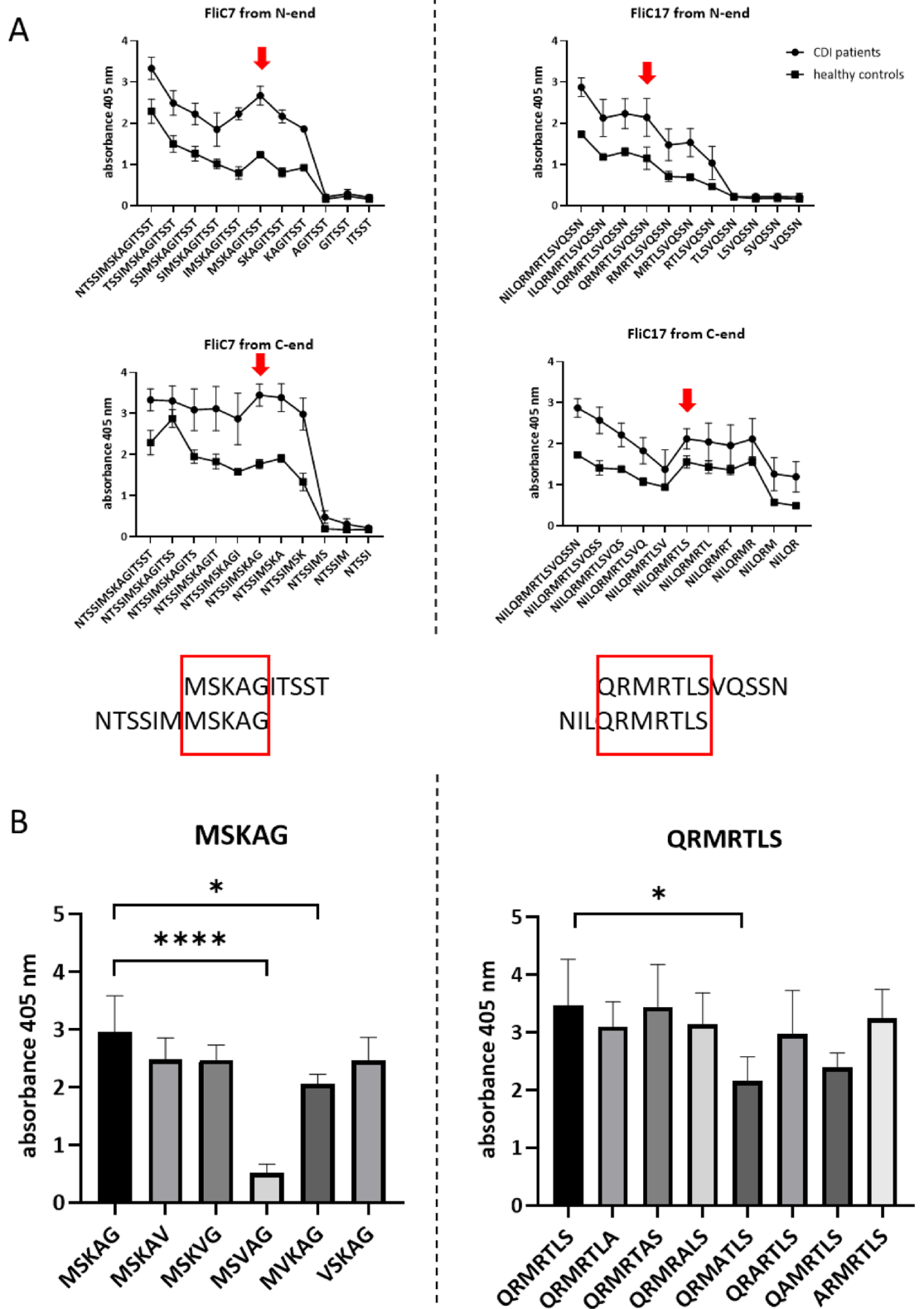
Figure 4. Mapping of *C. difficile* flagellin epitopes. (**A**) The analysis of the immunoreactivity of the library of truncated peptides of FliC7 and FliC17. Pin-bound ELISA was performed using pooled sera from *C. difficile*-infected patients and healthy volunteers (1:1000 dilution). Anti-human IgG conjugated with AP were used as secondary antibodies (1:10,000 dilution). Means with ± SD of three measurements are shown. (**B**) The "alanine walk" analysis of MSKAG and QRMRTLS epitopes. Pin-bound ELISA was performed using pooled *C. difficile*-infected patients' sera (1:1000 dilution). Anti-human IgG conjugated with AP were used as secondary antibodies (1:10,000 dilution). Means with ± SD of three measurements are shown. Data analyzed with one-way ANOVA; $*p < 0.0332$; $**p < 0.0021$; $***p < 0.0002$; $****p < 0.0001$.
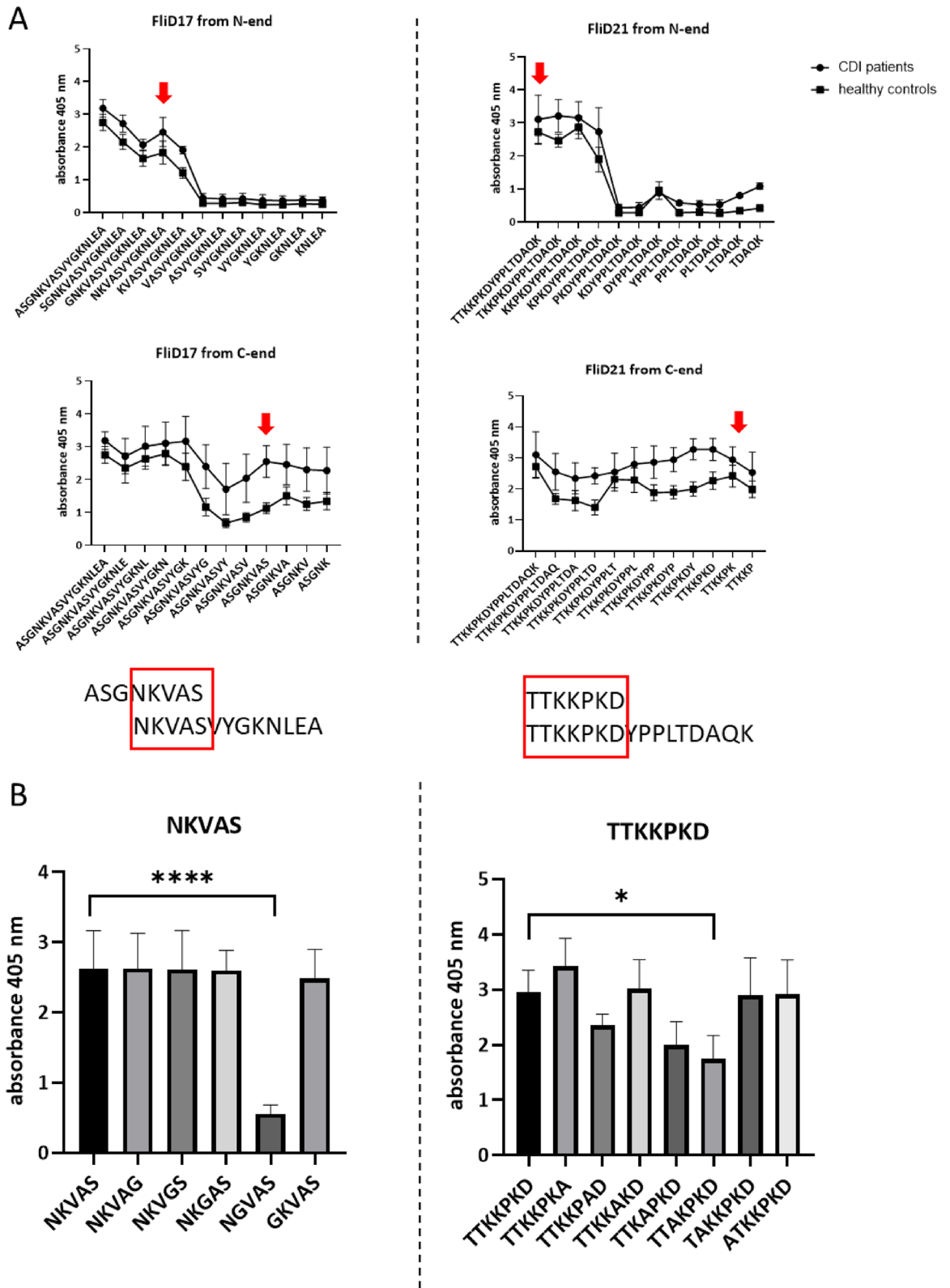
Figure 5. Mapping of *C. difficile* cap protein epitopes. (**A**) The analysis of the immunoreactivity of the library of truncated peptides of FliD17 and FliD21. Pin-bound ELISA was performed using pooled sera from *C. difficile*-infected patients and healthy volunteers (1:1000 dilution). Anti-human IgG conjugated with AP were used as secondary antibodies (1:10,000 dilution). Means with $\pm$ SD of three measurements are shown. (**B**) The "alanine walk" analysis of NKVAS and TTKKPKD epitopes. Pin-bound ELISA was performed using pooled *C. difficile*-infected patients' sera (1:1000 dilution). Anti-human IgG conjugated with AP were used as secondary antibodies (1:10,000 dilution). Means with $\pm$ SD of three measurements are shown. Data analyzed with one-way ANOVA; *$p < 0.0332$, **$p < 0.0021$, ***$p < 0.0002$, ****$p < 0.0001$.
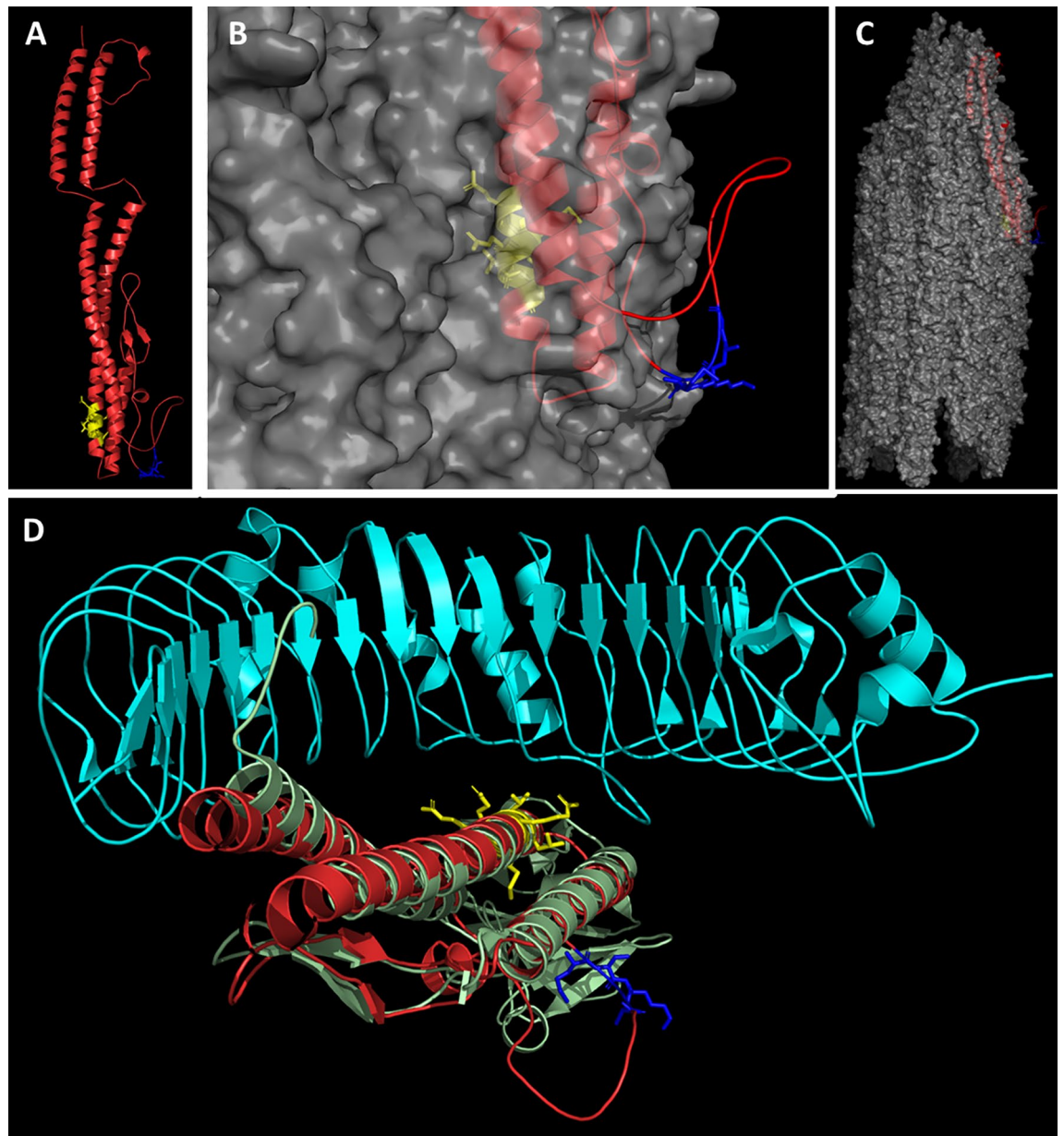
Figure 6. Modelling of *C. difficile* flagellin. (**A**) Secondary structure of FliC monomer. [205]MSKAG[209] epitope is indicated in blue. [117]QRMRTLS[123] epitope is indicated in yellow. (**B**, **C**) The FliC multimer which forms flagella body. (**D**) The interaction between TLR5 molecule (cyan) and FliC (green *Salmonella* FliC; red CD FliC). [117]QRMRTLS[123] epitope indicated in yellow in the center of the interaction.

In silico analysis of epitopes immunogenicity and autoimmunoreactivity. In silico analysis of the potential immunogenic properties of FliC and FliD epitopes by VaxiJen v.2.0 (threshold > 0.4) showed that three of the epitopes have antigenic properties: [205]MSKAG[209](G) 0.6749, [226]NKVAS[230](G) 1.5113 and [306]TTKKPKD[312] 0.7361. The [117]QRMRTLS[123] epitope did not pass the test (-0.137). We performed a search in the database of known epitopes and autoepitopes, none of the epitopes characterized by us was found. Moreover, none of these sequences were found in known human protein sequences.

Flagella modeling, epitope localization. We modeled the structure of FliC and FliD as monomers to localize the epitopes deteremined by us (Figs. 6A and 7A). Also, we modeled the structure of assembled flagella and flagellar cap hexamer to localize epitopes in the native bacterial protein structures (Figs. 6B,C and 7B). In the FliC protein one of the epitopes which is [205]MSKAG[209] is exposed to the environment and forms a loop which is usually easily accessible by antibodies. The [117]QRMRTLS[123] is buried inside flagella channel. We modelled the interaction between TLR5 and FliC based on crystallization data obtained for TLR5 and FliC from *Salmonella enterica* subsp. enterica serovar Typhimurium str. LT2. The results show that [117]QRMRTLS[123] is placed directly in the area of interaction between these two molecules.
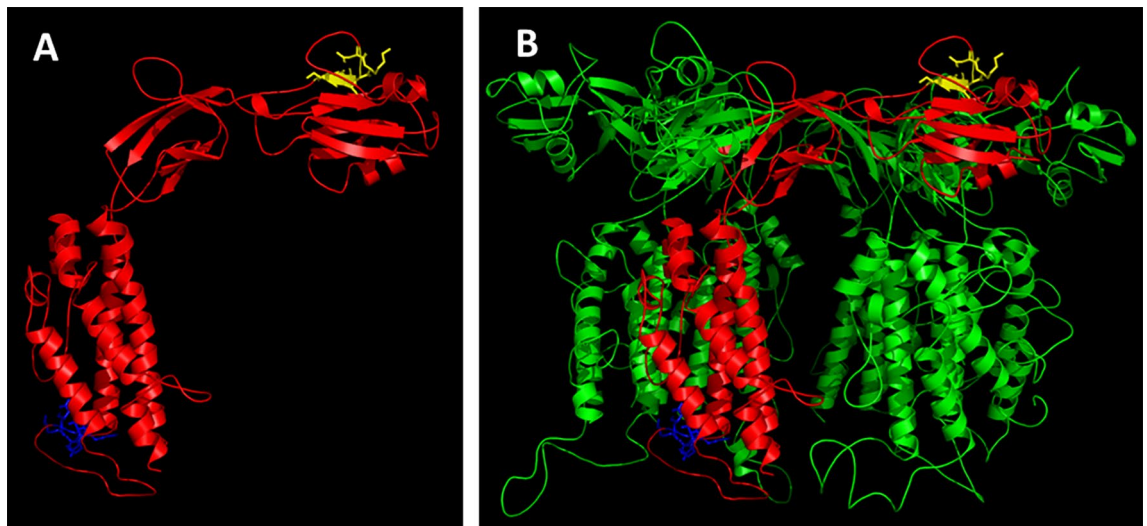
Figure 7. Modelling of *C. difficile* flagellar cap protein. (**A**) Secondary structure of FliD monomer. [226]NKVAS[230] epitope is indicated in yellow. [306]TTKKPKD[312] epitope is indicated in blue. (**B**) The FliD multimer that forms *C. difficile* flagellar cap.

In FliD, both epitopes are localized in structures other than alpha-helix which makes them more accessible for antibodies. FliD protein subunits form a hexamer and the two epitopes are localized on the two opposite sides of the hexamer. The analysis of FliD domains by Predict Protein server showed that [306]TTKKPKD[312] epitope is localized in the protein-binding domain.

Discussion

Flagella is a highly conserved virulence factor for many pathogens including *Pseudomonas aeruginosa*[34], *Salmonella* strains[35] or *C. difficle*[36]. Flagella is not only indispensable for pathogen motility but also takes part in attachment to the host cells by acting as an adhesion molecule; participates in biofilm formation; and interacts with hosts' immune cells, for example via Toll-like receptor 5 recognition[37]. Anti-flagella humoral response was shown to increase antibacterial function of host immune cells by phagocytosis stimulation[35]. Because of the above, flagella was already proposed as a suitable vaccine target[38] and some of these vaccines already reached clinical trials[39]. Flagellar proteins were also proposed as potential adjuvants since they elicit potent, mixed Th1/Th2 immune response[40,41]. Moreover, Bruxelle et al. proposed *C. difficile* FliC as a mucosal adjuvant in an anti-*C. difficile* vaccine[42]. The aim of our work was to identify the epitopes of both FliC and FliD proteins, which could be suitable for specific subunit vaccine development.

The humoral response of the host against *C. difficile* was extensively studied[8,9]. In the context of *C. difficile* flagella, it has been shown that the antibody levels against FliC and FliD were higher in the control group compared to the *C. difficile* positive group, which indicate their potentially protective role[8,9]. Compared to results of Péchiné et al.[8,9] we do not see such a relationship in our data, probably due to the fact that we have obtained sera at the earlier stage of infection and we measured the reactivity of specific anti-peptide IgG only instead of total anti-*C. difficile* immunoglobulins in patients' sera (Fig. 1). It is also worth mentioning that high level of antibodies reacting with *C. difficile* proteins might be a result of cross-reactivity. We observed a higher immunoreactivity of umbilical cord blood sera than healthy controls sera which was also noticed in our previous studies[6,7,26]. It is well known that cord blood serum has a significantly different composition in terms of antibodies than blood circulating in the system and can be a source of potentially protective antibodies[43]. Based on the high immunoreactivity of *C. difficile*-infected patients' sera and umbilical cord blood sera we were able to select several peptide sequences of FliC and FliD for further detailed analysis. Using bioinformatic and empirical methods, we were able to define the shortest peptide sequences involved in antibody binding and map the functional epitopes which consist of amino acids essential for epitope recognition by paratope (Figs. 4 and 5). We identified two epitopes in FliC protein, the main protein constituting the flagellar body. *C. difficile* FliC, like in many other Gram-positive bacteria, is composed of D0 and D1 domains required for TLR5 activation and flagellar body assembly[44]. [117]QRMRTLS[123] epitope identified by us using *C. difficile*-infected patients' blood sera (Fig. 4) is localized in a conserved hot spot responsible for binding and activation of TLR5, complementary to the LRR9 loop of TLR5 identified by Song et al.[45]. The first arginine in the [117]QRMRTLS[123] sequence (R118) is highly conserved in TLR5-activating flagellins and its substitution with alanine have significantly reduced TLR5 activation by ten fold[45]. Song et al. suggests that this specific arginine mutation changes shape and chemical complementarity between flagellin and TLR5 LRR9 loop. Likewise Song et al.[45], we observed the same effect in immunoreactivity of [117]QRMRTLS[123] after R118 substitution with alanine (Fig. 4B). However, Song et al. observed also a lower TLR5-binding effect in case of the mutation of L122 (two fold), similar effect was not shown by us in case of [117]QRMRTLS[123] immunoreactivity. Flagella activates TLR5 only in a monomeric form since sequences that are responsible for binding are buried in the assembled flagella[46]. The [117]QRMRTLS[123] epitope is localized in the

TLR5-binding region buried inside flagella body which was also shown by us in this study (Fig. 6B,D). Despite the fact that [117]QRMRTLS[123] epitope has interesting properties, i.e. it is located in the TLR5 binding region and is recognized by patients' antibodies, its use in a vaccine should be approached with caution as it shows a high level of cross-reactions (Fig. 2, [117]QRMRTLS[123] is a part of peptide 17, see also Supplementary Tab. S2).

The second of epitopes identified by us in FliC [205]MSKAG[209] is well exposed in the assembled flagella and easily accessible by antibodies. Moreover, it is a part of VALVNTSSIMSK sequence which was shown to be glycosylated[47]. Glycosylation may have a broad effect on epitopes ranging from its inactivation to being necessary for its recognition by antibodies[48]. However, even without the glycosylation, [205]MSKAG[209] epitope is very well recognized by *C. difficile*-infected patients' antibodies. The importance of [205]MSKAG[209] sequence for flagella functioning is unknown. Mapping of the functional epitope of [205]MSKAG[209] revealed that substitution of K207 with valine resulted in total loss of epitope immunoreactivity. It has been already shown by us that positively-charged lysine is critical for epitope-paratope interaction[26]. Since this epitope is characteristic for *C. difficile* and was positively verified by VaxiJen v2.0 as potentially immunogenic we plan to check its protective properties using mice model of CDI.

C. difficile FliD is a highly conserved protein that functions as a flagellar cap[49]. Antibodies specific for FliD are present in 87% of patients with diagnosed CDI which is a higher number than the one tested for FliC (21%) or *C. difficile* toxins A (60%) and B (25%)[8]. This means that FliD is highly immunogenic and might have a tremendous role in the effective humoral response against *C. difficile*. We mapped epitopes of *C. difficile* FliD using *C. difficile*-infected patients' sera. The 16-amino acid peptides designed for FliD mapping showed similar level of immunoreactivity as the ones synthesized for FliC (Fig. 1A vs B). Also, FliD-derived peptides were more immunoreactive with umbilical cord blood sera than healthy volunteers sera which indicates that the antibodies contained in the serum of umbilical cord blood are more concentrated, at least at the time of birth when the blood was sampled by us. This is in line with findings of de Voer et al. who tested 197 maternal and cord blood pairs and showed that umbilical cord blood antibody titers against protein antigens from selected pathogens were 1.6 times higher than those in maternal blood[50]. With our approach we were able to find epitopes of *C. difficile* FliD which are [226]NKVAS[230] and [306]TTKKPKD[312]. In each of them lysins were the most important amino acids for antibody binding (Fig. 5). Both passed the VaxiJen v2.0 test as immunogens and are well exposed in the FliD monomer and hexamer (Fig. 7). The analysis of FliD sequence with PredictProtein tool showed that [306]TTKKPKD[312] is localized in a protein-binding region. The *C.* difficile FliD hexamer forms table-like structure probably similar to the one crystalized for *Pseudomonas aeruginosa*[51]. According to structures shown by Postel et al. and our model of *C. difficile* FliD [226]NKVAS[230] is located in the 'head' region of FliD and [306]TTKKPKD[312] in the 'leg' region of this protein which is highly flexible and probably involved in protein oligomerization. The importance of these regions localization and their functions in the context of functional flagella assembly will be further investigated.

Another aspect of our work is the potential cross-reactivity of FliC and FliD peptides with host proteins or molecules present in its microbiome. Current solutions are based on the comparison of amino acid sequences. However, given that the interaction between epitope and paratope takes place at the atomic level and not amino acid residue level[52], a bioinformatic-based approach may not always detect potential unwanted cross-reactions. Cross-reactivity analysis conducted in this study between *C. difficile* FliC and FliD peptides and sera obtained from rabbits immunized with flagellated bacteria strains showed that even a very small sequence homology can give potentially unwanted cross-reactions, and in turn, peptides with high homology to the antigen used for immunizations do not have to be highly immunoreactive (Figs. 2 and 3). These disparity may be due to the fact that the same antigens can immunize various organisms differently (human *vs* rabbit)[53]. Therefore, given that most studies are conducted on mice and then on humans, these differences should be taken into account. However, in some cases cross-reactivity might have good effects. Studies show that vaccination with pneumococcal vaccines might protect against SARS-CoV-2 infection and death[54]. Pvs48/45 antigens from *Plasmodium vivax* and *P. falciparum* share 55% identity and showed high cross-reactivity and cross-boosting properties[55]. This research gives hope for new vaccine targeting transmission of both species of malaria parasites which are responsible for the majority of malaria infections in the world. However, there are numerous reports linking molecular mimicry, resulting cross-reactions and autoimmunity with vaccination[56]. Solving this problem is particularly important in the era of increasing vaccine hesitancy to which the vaccine safety concern is one of the strongest argument. We propose that not only pathogen-unique peptides/structures should be considered to use as a vaccine antigen but also they should be thoroughly tested empirically in terms of possible cross- and auto-immunoreactivity. This is particularly important in the context of the use of the entire flagella protein as an adjuvant.

Within this paper, using bioinformatics methodology together with the analysis of patient's sera immunoreactivity, we identified four epitopes of *C. difficile* flagella proteins. Our approach is in line with the currently recommended algorithm of searching for new candidates for vaccine antigens that advises the use of amino acid sequences with very low homology to self-antigens[57], and we even extend our investigations with the analysis of cross-reactions that may occur between seemingly completely different sequences. We showed that [117]QRMRTLS[123] sequence identified by us in *C. difficile* flagellin is not only TLR5-binding and activating, as previously shown, but forms an epitope recognized by *C. difficile*-infected patients' antibodies. [205]MSKAG[209] is a *C. difficile*-unique, immunogenic sequence that forms an exposed epitope on the polymerized flagella structure which makes it suitable vaccine antigen. [226]NKVAS[230] and [306]TTKKPKD[312] are two FliD epitopes, which are well exposed and protective according to VaxiJen analysis. Further, their potential protective properties will be verified by us in a mouse model of CDI.

Data availability
The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Guh, Y. & Kutty, K. Clostridioides difficile infection. *Ann. Intern. Med.* **169**, 49–64 (2018).
2. Li, H. *et al.* Antibiotic resistance of clinical isolates of clostridioides difficile in china and its association with geographical regions and patient age. *Anaerobe* **60**, 102094. https://doi.org/10.1016/j.anaerobe.2019.102094 (2019).
3. Roo, A. C. D. & Regenbogen, S. E. Clostridium difficile infection: an epidemiology update. *Clin. Colon Rectal Surg.* **33**, 49–57. https://doi.org/10.1055/s-0040-1701229 (2020).
4. Péchiné, S., Bruxelle, J. F., Janoir, C. & Collignon, A. Targeting clostridium difficile surface components to develop immunotherapeutic strategies against clostridium difficile infection. *Front. Microbiol.* https://doi.org/10.3389/fmicb.2018.01009 (2018).
5. Wright, A., Drudy, D., Kyne, L., Brown, K. & Fairweather, N. F. Immunoreactive cell wall proteins of clostridium difficile identified by human sera. *J. Med. Microbiol.* **57**, 750–756. https://doi.org/10.1099/jmm.0.47532-0 (2008).
6. Razim, A. *et al.* Mapping epitopes of a novel peptidoglycan cross-linking enzyme Cwp22 recognized by human sera obtained from patients with clostridioides difficile infection and cord blood. *Microorganisms* **7**, 565. https://doi.org/10.3390/microorganisms7110565 (2019).
7. Razim, A. *et al.* Epitopes identified in GAPDH from clostridium difficile recognized as common antigens with potential autoimmunizing properties. *Sci. Rep.* **8**, 13946. https://doi.org/10.1038/s41598-018-32193-9 (2018).
8. Pechine, S. Immunological properties of surface proteins of clostridium difficile. *J. Med. Microbiol.* **54**, 193–196. https://doi.org/10.1099/jmm.0.45800-0 (2005).
9. Pechine, S., Janoir, C. & Collignon, A. Variability of clostridium difficile surface proteins and specific serum antibody response in patients with clostridium difficile-associated disease. *J. Clin. Microbiol.* **43**, 5018–5025. https://doi.org/10.1128/JCM.43.10.5018-5025.2005 (2005).
10. Garcia, M. *et al.* Pseudomonas aeruginosa flagellum is critical for invasion, cutaneous persistence and induction of inflammatory response of skin epidermis. *Virulence* **9**, 1163–1175. https://doi.org/10.1080/21505594.2018.1480830 (2018).
11. Gu, H. Role of flagella in the pathogenesis of helicobacter pylori. *Curr. Microbiol.* **74**, 863. https://doi.org/10.1007/s00284-017-1256-4 (2017).
12. de Barbosa, F. O. *et al.* Contribution of flagella and motility to gut colonisation and pathogenicity of salmonella enteritidis in the chicken. *Braz. J. Microbiol.* **48**, 754–759. https://doi.org/10.1016/j.bjm.2017.01.012 (2017).
13. Cui, B. *et al.* Flagellin as a vaccine adjuvant. *Expert Rev. Vaccines* **17**, 335–349. https://doi.org/10.1080/14760584.2018.1457443 (2018).
14. Tasteyre, A., Barc, M.-C., Collignon, A., Boureau, H. & Karjalainen, T. Role of FliC and FliD flagellar proteins of clostridium difficile in adherence and gut colonization. *Infect. Immun.* **69**, 7937–7940. https://doi.org/10.1128/IAI.69.12.7937-7940.2001 (2001).
15. Đapa, T. *et al.* Multiple factors modulate biofilm formation by the anaerobic pathogen clostridium difficile. *J. Bacteriol.* **195**, 545–555. https://doi.org/10.1128/JB.01980-12 (2013).
16. Batah, J. *et al.* Clostridium difficile flagella induce a pro-inflammatory response in intestinal epithelium of mice in cooperation with toxins. *Sci. Rep.* **7**, 3256. https://doi.org/10.1038/s41598-017-03621-z (2017).
17. Ghose, C. *et al.* Immunogenicity and protective efficacy of recombinant *clostridium difficile* flagellar protein FliC. *Emerg. Microbes Infect.* **5**, e8. https://doi.org/10.1038/emi.2016.8 (2016).
18. Aptekorz, M. *et al.* Occurrence of clostridium difficile ribotype 027 in hospitals of Silesia, Poland. *Anaerobe* **45**, 106–113. https://doi.org/10.1016/j.anaerobe.2017.02.002 (2017).
19. Gamian, A., Romanowska, A. & Romanowska, E. Immunochemical studies on sialic acid-containing lipopolysaccharides from enterobacterial species. *FEMS Microbiol. Immunol.* **4**, 323–328. https://doi.org/10.1111/j.1574-6968.1992.tb05012.x (1992).
20. Gamian, A. & Romanowska, E. The core structure of Shigella sonnei lipopolysaccharide and the linkage between o-specific polysaccharide and the core region. *Eur. J. Biochem.* **129**, 105–109. https://doi.org/10.1111/j.1432-1033.1982.tb07027.x (1982).
21. Coordinators, N. R. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **44**, D7–D19. https://doi.org/10.1093/nar/gkv1290 (2016).
22. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641. https://doi.org/10.1093/nar/gkz268 (2019).
23. Lian, Y., Ge, M. & Pan, X.-M. EPMLR: sequence-based linear B-cell epitope prediction method using multiple linear regression. *BMC Bioinform.* **15**, 10. https://doi.org/10.1186/s12859-014-0414-y (2014).
24. Yao, B., Zhang, L., Liang, S. & Zhang, C. SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. *PLoS ONE* **7**, e45152. https://doi.org/10.1371/journal.pone.0045152 (2012).
25. Carter, J. M. Epitope mapping of a protein using the geysen (PEPSCAN) procedure. *Pept. Anal. Protoc.* **36**, 207–224 (2014).
26. Pacyga, K. *et al.* The bioinformatic and in vitro studies of clostridioides difficile aminopeptidase M24 revealed the immunoreactive KKGIK peptide. *Cells* **9**, 1146. https://doi.org/10.3390/cells9051146 (2020).
27. Blüthner, M., Koch, J. & Mahler, M. Mutational analysis and structure predictions. In *Peptide Arrays on Membrane Supports: Synthesis and Applications* (eds Koch, J. & Mahler, M.) 123–139 (Springer, Berlin, 2002).
28. Doytchinova, I. A. & Flower, D. R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* **8**, 4. https://doi.org/10.1186/1471-2105-8-4 (2007).
29. Vita, R. *et al.* The immune epitope database (IEDB) 30. *Nucleic Acids Res.* **43**, D405–D412. https://doi.org/10.1093/nar/gku938 (2015).
30. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858. https://doi.org/10.1038/nprot.2015.053 (2015).
31. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303. https://doi.org/10.1093/nar/gky427 (2018).
32. DeLano, W.L. *Pymol, DeLano Scientific, San Carlos, CA, 700*; 2006.
33. Yachdav, G. *et al.* PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res.* **42**, W337–W343. https://doi.org/10.1093/nar/gku366 (2014).
34. Campodónico, V. L. *et al.* Evaluation of flagella and flagellin of pseudomonas aeruginosa as vaccines. *Infect. Immun.* **78**, 746–755. https://doi.org/10.1128/IAI.00806-09 (2010).
35. Goh, Y. S., Armour, K. L., Clark, M. R., Grant, A. J. & Mastroeni, P. Igg subclasses targeting the flagella of salmonella enterica serovar typhimurium can mediate phagocytosis and bacterial killing. *J. Vaccines Vaccin.* **7**, 10. https://doi.org/10.4172/2157-7560.1000322 (2016).
36. Stevenson, E., Minton, N. P. & Kuehne, S. A. The role of flagella in clostridium difficile pathogenicity. *Trends Microbiol.* **23**, 275–282. https://doi.org/10.1016/j.tim.2015.01.004 (2015).
37. Duan, Q., Zhou, M., Zhu, L. & Zhu, G. Flagella and bacterial pathogenicity. *J. Basic Microbiol.* **53**, 1–8. https://doi.org/10.1002/jobm.201100335 (2013).
38. Skene, C., Young, A., Every, A. & Sutton, P. Helicobacter pylori flagella: antigenic profile and protective immunity. *FEMS Immunol. Med. Microbiol.* **50**, 249–256. https://doi.org/10.1111/j.1574-695X.2007.00263.x (2007).

39. Döring, G., Meisner, C. & Stern, M., Group¶ for the F.V.T.S. A double-blind randomized placebo-controlled phase III study of a pseudomonas aeruginosa flagella vaccine in cystic fibrosis patients. *Proc. Natl. Acad. Sci.* **104**, 11020–11025. https://doi.org/10.1073/pnas.0702403104 (2007).
40. Huleatt, J. W. *et al.* Vaccination with recombinant fusion proteins incorporating toll-like receptor ligands induces rapid cellular and humoral immunity. *Vaccine* **25**, 763–775. https://doi.org/10.1016/j.vaccine.2006.08.013 (2007).
41. Gries, C. M., Mohan, R. R., Morikis, D. & Lo, D. D. Crosslinked flagella as a stabilized vaccine adjuvant scaffold. *BMC Biotechnol.* **19**, 48. https://doi.org/10.1186/s12896-019-0545-3 (2019).
42. Bruxelle, J.-F. *et al.* Clostridium difficile flagellin FliC: evaluation as adjuvant and use in a mucosal vaccine against clostridium difficile. *PLoS ONE* **12**, e0187212. https://doi.org/10.1371/journal.pone.0187212 (2017).
43. Brzychczy-Wloch, M. *et al.* Identification of high immunoreactive proteins from *Streptococcus Agalactiae* isolates recognized by human serum antibodies. *FEMS Microbiol. Lett.* https://doi.org/10.1111/1574-6968.12292 (2013).
44. Zeigler, D. R. The genome sequence of Bacillus Subtilis Subsp Spizizenii W23: insights into speciation within the B. Subtilis complex and into the history of B. Subtilis genetics. *Microbiology* **157**, 2033–2041. https://doi.org/10.1099/mic.0.048520-0 (2011).
45. Song, W. S., Jeon, Y. J., Namgung, B., Hong, M. & Yoon, S. A conserved tlr5 binding and activation hot spot on flagellin. *Sci. Rep.* **7**, 40878. https://doi.org/10.1038/srep40878 (2017).
46. Smith, K. D. *et al.* Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. *Nat. Immunol.* **4**, 1247–1253. https://doi.org/10.1038/ni1011 (2003).
47. Bouché, L. *et al.* The type B flagellin of hypervirulent clostridium difficile is modified with novel sulfonated peptidylamido-glycans. *J. Biol. Chem.* **291**, 25439–25449. https://doi.org/10.1074/jbc.M116.749481 (2016).
48. Lisowska, E. The role of glycosylation in protein antigenic properties. *Cell. Mol. Life Sci. CMLS* **59**, 445–455. https://doi.org/10.1007/s00018-002-8437-3 (2002).
49. Tasteyre, A. *et al.* Molecular characterization of FliD gene encoding flagellar cap and its expression among clostridium difficilei-solates from different serogroups. *J. Clin. Microbiol.* **39**, 1178–1183. https://doi.org/10.1128/JCM.39.3.1178-1183.2001 (2001).
50. de Voer, R. M. *et al.* Seroprevalence and placental transportation of maternal antibodies specific for neisseria meningitidis sero-group C, haemophilus influenzae type B, diphtheria, tetanus, and pertussis. *Clin. Infect. Dis.* **49**, 58–64. https://doi.org/10.1086/599347 (2009).
51. Postel, S. *et al.* Bacterial flagellar capping proteins adopt diverse oligomeric states. *Elife* **5**, e18857. https://doi.org/10.7554/eLife.18857 (2016).
52. Rauffer, N., Zeder-Lutz, G., Wenger, R., Van Regenmortel, M. H. V. & Altschuh, D. Structure-activity relationships for the interaction between cyclosporin a derivatives and the fab fragment of a monoclonal antibody. *Mol. Immunol.* **31**, 913–922. https://doi.org/10.1016/0161-5890(94)90011-6 (1994).
53. Jarząb, A. *et al.* Shigella flexneri 3a outer membrane protein C epitope is recognized by human umbilical cord sera and associated with protective activity. *PLoS ONE* **8**, e70539. https://doi.org/10.1371/journal.pone.0070539 (2013).
54. Root-Bernstein, R. Possible cross-reactivity between SARS-CoV-2 proteins, CRM197 and proteins in pneumococcal vaccines may protect against symptomatic SARS-CoV-2 disease and death. *Vaccines* **8**, 559. https://doi.org/10.3390/vaccines8040559 (2020).
55. Cao, Y., Bansal, G. P., Merino, K. & Kumar, N. Immunological cross-reactivity between malaria vaccine target antigen P48/45 in plasmodium vivax and P. falciparum and cross-boosting of immune responses. *PLoS ONE* **11**, e0158212. https://doi.org/10.1371/journal.pone.0158212 (2016).
56. Segal, Y. & Shoenfeld, Y. Vaccine-induced autoimmunity: the role of molecular mimicry and immune crossreaction. *Cell. Mol. Immunol.* **15**, 586–594. https://doi.org/10.1038/cmi.2017.151 (2018).
57. Kanduc, D. "Self-nonself" peptides in the design of vaccines. *Curr. Pharm. Des.* **15**, 3283–3289 (2009).

Author contributions

R.A. planned work, analyzed results, performed immunoreactivity analysis, wrote the first draft and edited manuscript, P.K. prepared peptides, analyzed their immunoreactivity, N.P. performed cross-reactivity analysis, M.D. analyzed epitope localization and prepared protein models, S.A. prepared patient serum library, G.A. obtained funding, edited the manuscript, G.S. planned work and analyzed data, edited manuscript.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-89488-7.

Correspondence and requests for materials should be addressed to A.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.