CrossMark

ORIGINAL ARTICLE

Surface proteome mining for identification of potential vaccine candidates against *Campylobacter jejuni*: an in silico approach

Kusum Mehla[1] · Jayashree Ramana[1]

Abstract *Campylobacter jejuni* remains a major cause of human gastroenteritis with estimated annual incidence rate of 450 million infections worldwide. *C. jejuni* is a major burden to public health in both socioeconomically developing and industrialized nations. Virulence determinants involved in *C. jejuni* pathogenesis are multifactorial in nature and not yet fully understood. Despite the completion of the first *C. jejuni* genome project in 2000, there are currently no vaccines in the market against this pathogen. Traditional vaccinology approach is an arduous and time extensive task. Omics techniques coupled with sequencing data have engaged researcher's attention to reduce the time and resources applied in the process of vaccine development. Recently, there has been remarkable increase in development of in silico analysis tools for efficiently mining biological information obscured in the genome. In silico approaches have been crucial for combating infectious diseases by accelerating the pace of vaccine development. This study employed a range of bioinformatics approaches for proteome scale identification of peptide vaccine candidates. Whole proteome of *C. jejuni* was investigated for varied properties like antigenicity, allergenicity, major histocompatibility class (MHC)–peptide interaction, immune cell processivity, HLA distribution, conservancy, and population coverage. Predicted epitopes were further tested for binding in MHC groove using computational docking studies. The predicted epitopes were conserved; covered more than 80 % of the world population and were presented by MHC-I supertypes. We conclude by underscoring that the epitopes predicted are believed to expedite the development of successful vaccines to control or prevent *C. jejuni* infections albeit the results need to be experimentally validated.

Keywords *C. jejuni* · Vaccine candidates · Reverse vaccinology · Antigenicity · Allergenicity · Docking

Introduction

Campylobacter jejuni are helical-shaped, non-spore forming, microaerophilic gram-negative bacteria and are a major cause of bacterial campylobacteriosis worldwide (Friedman et al. 2000). The *Campylobacter* spp. was considered as zoonotic pathogen until isolation of *C. jejuni* was accomplished from human feces in 1968 (Dekeyser et al. 1972). Since its discovery in the 1970s, *C. jejuni* remains the most frequent cause of infectious diarrhea affecting over 450 million people every year throughout the world, attributing to a large economic burden (Friedman et al. 2000). The first *C. jejuni* (strain NCTC11168) genome was sequenced in 2000 with 94.3 % of the genome coding for proteins (Parkhill et al. 2000). *C. jejuni* pathogenesis mechanisms are poorly understood as virulence determinants appear to be multifactorial in nature such as chemotaxis, motility, toxins, flagella, invasion and adherence, and surface polysaccharide structures (Ketley 1997). Antibiotic therapy traditionally involves treatment with erythromycin and ciprofloxacin, but many reports have witnessed resistance of *C. jejuni* to different antibiotics such as tetracycline, kanamycin, chloramphenicol, erythromycin, and ciprofloxacin (Alfredson and Korolik 2007, Thakur et al. 2010). Due to irrational use of antibiotics, antibiotic resistance has escalated posing a challenge to current treatment regimens. Thus, there is a pressing need to develop alternative treatments.

✉ Jayashree Ramana
jayashree.ramana@juit.ac.in

[1] Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Solan PIN-173234, Himachal Pradesh, India

Springer

Vaccination has proven to be a cost effective, safe, and efficient solution to combat infectious diseases like meningococcal, diphtheria, tetanus, poliomyelitis, pertussis, measles, mumps, and rubella in human health care (Moriel et al. 2008). The traditional approach to subunit vaccine development has negative aspects involving time and labor intensive nature, failure in cases where microorganism cannot be cultured or obtained in sufficient amounts (Rinaudo et al. 2009). For limiting increasing antibiotic resistance and increasing number of human infections, developing vaccines against *C. jejuni* is both indispensable and attractive. Some mutants of *C. jejuni* with defects in pili or invasion biosynthesis are being evaluated for their protective efficacy in animal models. Flagellin and adhesin proteins have been suggested as potential subunit-based vaccine candidates such as a recombinant-truncated flagellin protein (rFla-MBP) conferred 60 % protection in a ferret model of diarrhea. Several killed whole-cell (WC) and heat-labile toxin (LT) adjuvanted vaccines are under development (O'Ryan et al. 2015, Albert 2014). In one such example, killed *Campylobacter* whole-cell (CWC) organism adjuvanted with heat-labile enterotoxin (LT) of *Escherichia coli* showed protection against intestinal colonization in mice and rabbits. However, currently, there are no approved vaccines available to treat *Campylobacter*-associated illness. Sequencing the genome of many *Campylobacter* strains together with development of omics techniques and advanced bioinformatics approaches significantly improve the process of candidate epitope identification minimizing the arduous peptide screening task for immunobiological properties. The present study has employed a range of computational approaches to investigate the entire proteome of *C. jejuni* for identification of B- and T-cell epitopes as potential vaccine candidates. This study has important repercussion for selection of vaccine candidates, a critical step in vaccine development.

Methods

Retrieving non-homologous proteins from pathogen whole proteome

As described in the workflow diagram (Fig. 1), the complete proteome of the *C. jejuni* O:2 (strain NCTC 11168) encoding 1623 proteins was retrieved from Uniprot (Proteome ID UP000000799). Proteins non-homologous to host from pathogen proteome were retrieved using a two-step filtration procedure. In the first step, sequences with length less than 100 amino acids (aa) were filtered based on the fact that the average protein length in bacteria is 267 aa (Brocchieri and Karlin 2005). Consequently proteins with length less than 100 aa would probably not code for any protein. In the next filtration step, sequences were further screened out based on homology with the host (*Homo sapiens*) proteome at an *e* value cutoff of

0.05. In the BLASTp search, proteins which showed no hits below *e* value inclusion threshold were selected as non-homologous pathogen proteins.

Antigenicity and transmembrane prediction

To predict antigenic sequences, these non-homologous pathogen proteins were subjected to VaxiJen server (Doytchinova and Flower 2007), which is based on auto cross covariance (ACC) transformation of protein sequences into uniform vectors of principal amino acid properties with a threshold value of 0.7. The sequences with antigenicity value above threshold were subjected to PSORTb version 3.0 to retrieve outer membrane localized proteins. PSORTb utilizes a Bayesian network model to calculate associated probability for five localization sites viz. cytoplasmic, inner membrane, periplasmic, outer membrane, and extracellular with a default probability value (*p* value) of 7.5 (Yu et al. 2010).

T-cell epitope prediction

NetCTL 1.2 Server was used to predict cytotoxic T lymphocyte (CTL) epitopes from the antigenic sequences localized in outer membrane, at a threshold value of 0.75 to maintain high sensitivity and specificity levels, and the prediction was restricted to 12 major histocompatibility class I (MHC-I) supertypes. NetCTL is an artificial neural network (ANN) and weight matrix-based tool combining the prediction of peptide MHC-I binding, proteasomal C terminal cleavage, and TAP transport efficiency (Larsen et al. 2007). The CTL epitopes generated from NetCTL were assessed for their allergenicity by subjecting them to AllerHunter program which is based on support vector machine (SVM) and pair-wise sequence similarity (Muh et al. 2009). A threshold value of 0.06 was specified for prediction of cross-reactive allergen.

An Immune Epitope Database (IEDB) tool based on combined predictors of proteasomal processing, TAP transport, and MHC binding was used for predictions of antigen processing through MHC-I (Tenzer et al. 2005). IEDB is the most inclusive database of experimentally characterized B- and T-cell epitopes. The stabilized matrix-based method (SMM) which can model the sequence specificity of quantifiable biological processes (Peters and Sette 2005) was employed to compute inhibitory concentration (IC50) values of peptide binding to MHCI molecules. In conjunction with the IEDB tool, MHCPred which uses a partial least squares-based multivariate statistical approach (Guan et al. 2003) was used for prediction of both MHC-I and MHC-II binders of the predicted peptides. The alleles with binding affinity IC50 value less than 500 nM from both the servers were considered as efficient peptide binders.

Fig. 1 Schematic representation of the protocols used for epitope identification
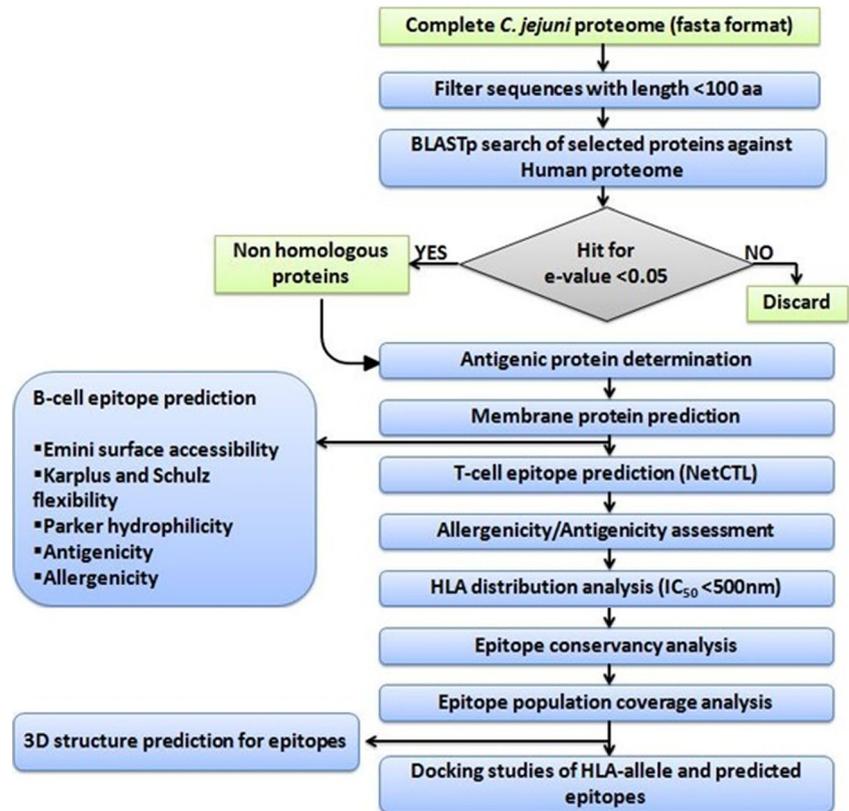
Epitope conservancy and HLA distribution analysis

For each identified peptide, the conservancy was predicted using the IEDB tool (Bui et al. 2007). The degree of conservation of each peptide was calculated as the fraction of protein sequences of different strains retrieved from UniProt that match the aligned peptide sequence above a defined identity level. An IEDB-based tool for human population coverage analysis (Bui et al. 2006) was used to study the distribution of human HLA alleles among the predicted epitopes. The predicted peptides with their corresponding MHC-I and MHC-II alleles were submitted with default parameter settings (the final set containing frequencies of 3245 alleles for 16 geographical areas, 21 ethnicities, and 115 countries). The predictions were made using the latest dataset from the Allele Frequency Net Database (AFND) (Gonzalez-Galarza et al. 2011).

Molecular docking studies of HLA-epitope

Designing epitope 3D structure

To study the molecular interactions between the predicted T-cell epitopes (YIQDNFNFY and NTDQAQGTV) and HLA molecules, PEP-FOLD based on a hidden Markov model-derived structural alphabet (SA) (Thevenet et al. 2012) was used to predict the 3D structure of the peptide. PEP-FOLD generated five models for input peptide sequence. The best model was selected for docking studies.

Docking

To validate our results, we performed a docking study of HLA-A*11:01 and selected epitope using Hex, the first Fourier transform (FFT)-based protein docking server (Macindoe et al. 2010). The crystal structure of HLA-A*11:01 in complex with sars nucleocapsid peptide (PDB Id: 1X7Q) was simplified to HLA-A*11:01, prepared by adding hydrogen atoms. Finally, the docking was carried out in Hex using prepared HLA-A*11:01 and PEP-FOLD predicted epitopes as starting structures. The parameters were set to default except for correlation type which uses both shape and electrostatics criteria for docking calculations. The best conformation was selected based on the Etotal (binding affinity) value, and complexes and interactions were visualized in PyMOL molecular graphics package (Schrodinger 2010) and Ligplot, respectively (Laskowski and Swindells 2011).

B-cell epitope identification

BCPred (El-Manzalawy et al. 2008) and AAP (Chen et al. 2007) methods at BCPred server, both of which use SVM-based classifiers, were utilized with an aim to identify potential antigens which can interact with B lymphocytes. Tools

from IEDB were employed to find the B-cell epitopes and further screen out the potential epitopes. Emini surface accessibility prediction (Emini et al. 1985), Karplus and Schulz flexibility prediction (Karplus and Schulz, 1985), and Parker hydrophilicity prediction (Parker et al. 1986) programs were used from IEDB. The regions common to predictions from both BCPred server and IEDB tools were considered as potential B-cell epitopes. These epitopes were further filtered based on allergenicity and antigenicity criteria using AllerHunter and VaxiJen, respectively.

Results

Retrieving non-homologous proteins from pathogen whole proteome

C. jejuni O:2 (strain NCTC 11168) whole proteome encodes 1623 proteins. After filtering out protein sequences on length criteria, we were left with 1500 proteins. We subjected the rest of the protein sequences to a homology search against human proteome database using BLASTP search from a standalone blast suite and retrieved a total of 210 pathogen proteins which were non-homologous to humans. Identifying proteins non-homologous to humans is essential as it excludes the possibility of the peptide vaccine targeting hosts enzymes, thus avoiding adverse effects on humans (Butt et al. 2012). Besides, self-peptides can mount an autoimmune response in the host.

Antigenicity and transmembrane prediction

The VaxiJen server used to assess the antigenicity of the protein sequences predicted 157 proteins as antigenic above a threshold of 0.7 which were further analyzed for their cellular location, and it revealed that 24 proteins were localized in outer membrane. Identification of outer membrane proteins is critical for reliable and rapid identification of vaccine candidates as many of the vaccines that trigger immune responses appeared to be secreted toxins or surface exposed molecules (Doro et al. 2009). Outer membrane localized proteins were further analyzed for vaccine candidate identification.

T-cell epitope prediction

NetCTL predicted T-cell epitopes from each sequence against MHC supertypes. Twenty-eight epitopes with their combinatorial score above threshold 2 were selected from the outer membrane localized antigenic proteins. These epitopes were further assessed by AllerHunter for allergic cross-reactivity and by VaxiJen for antigenicity. This step identified four epitopes as potential T-cell epitopes (Table 1). For each epitope, SMM-based IEDB MHC-I processing prediction tool retrieved the MHC-I alleles with IC50 value less than 500 nM

which were potential epitope binders. MHCPred predictions of MHC-I and MHC-II alleles as efficient epitope binders were taken together with IEDB tool predictions to generate a final list of potential binders for each epitope. The results are summarized in Table 2.

Epitope conservancy and HLA distribution analysis

For each predicted epitope, conservancy was determined using IEDB conservancy tool, and the results are shown in Table 2. Epitope NTDQAQGTV was 75 % conserved at identity of >60 %, while YIQDNFNFY was 50 % conserved at 100 % identity. Conservancy results for other epitopes (RSDEAQTNY and KSDEEMEKY) were not alluring. Due to scarcity of sequence data in UniProt database, conservancy results do not portray factual depiction of epitope conservancy. Population coverage analysis was then performed for epitopes NTDQAQGTV and YIQDNFNFY along with their associated MHC-I and MHC-II alleles as input to IEDB population coverage analysis tool. As shown in Table 3, immune response elicitation of the 81.07 and 85.27 % world population was covered by the epitopes NTDQAQGTV and YIQDNFNFY, respectively. Maximum coverage 85.99 % for epitope NTDQAQGTV was in Europe area followed by 85.53, 84.25, and 80.44 % in the population of South Africa, South Asia, and North Africa, respectively. For epitope YIQDNFNFY, maximum coverage 90.81 % was in Europe area followed by South Asia, North America, and Northeast Asia with coverage 85.70, 84.08, and 82.80 %, respectively.

Molecular docking studies of HLA-epitope

PEP-FOLD generates peptide structures by performing a series of simulations based on structural alphabet (SA) profiles derived from amino acid sequences. PEP-FOLD then returns the representative configuration for the input epitope based on energy and population parameters. Using Hex, different conformations of the predicted epitopes bound in MHC cleft were generated. Hex correlates molecules to 3D parametric functions such as electrostatic charge, surface shapes, and potential dissemination which define electrostatic and van der Walls interactions. The best conformation was then selected based on binding affinity scores which is dependent on such interactions. The docked complexes were visualized in Pymol as shown in Fig. 2. HLA-A*11:01 binds with epitopes NTDQAQGTV and YIQDNFNFY with binding energies −386.53 and −350.09 kcal/mol, respectively. Figure 3 represents the interactions involved in HLA-A*11:01 binding with predicted epitopes. Epitope NTDQAQGTV interacts with HLA-A*11:01 through hydrogen bonds with Tyr 27 and van der Walls interactions with MHC residues Ser 4, Arg 6, Phe 8, Asp 30, Gln 96, Met 98, Tyr 113, Ala 211, Glu 212, Thr 233, and Phe 241. MHC interacts with epitope YIQDNFNFY

Table 1 Most probable predicted epitopes selected on the basis of their NetCTL (MHC binding, proteasomal processing, and TAP transport), AllerHunter (allergic cross-reactivity) and VaxiJen (antigenicity) score

Protein_Id	Protein_Name	Gene_Name	Epitope	NetCTL score	AllerHunter score	VaxiJen score
Q0PBW1	Putative TonB-dependent outer membrane receptor	Cj0178	NTDQAQGTV	2.0440	0.04	1.1875
Q0P9M1	Outer membrane component of efflux system (multidrug efflux system cmeDEF)	cmeD	YIQDNFNFY	2.2174	0.05	0.6824
Q0P9S7	Putative outer membrane protein	Cj0975	RSDEAQTNY	3.3776	0.06	1.5564
Q0P826	Putative conserved protein (uncharacterized protein)	Cj1602	KSDEEMEKY	3.3309	0.06	1.1041

through hydrogen-bonded interactions with Asp 29 and Asp 30. Asp 30 forms two hydrogen bonds with Tyr 1 and Ile 2 in the epitopic sequence having bond lengths 2.86 and 2.87 Å, and Asp 29 is hydrogen bonded to Tyr 1 with a bond length of 2.57 Å. Epitope YIQDNFNFY is bound in MHC cleft due to hydrophobic interactions with Arg 6, Phe 8, Asp 102, Pro 210, Ala 211, Glu 212, Glu 232, Thr 233, Arg 234, Pro 235, Lys 243, and Phe 241. Involvement of common residues in interaction with different peptides suggests the crucial role of Arg 6, Phe 8, Ala 211, Glu 212, and Phe 241 MHC residues in MHC-peptide binding.

B-cell epitope identification

As per the criteria set for prediction of B-cell epitopes, Table 4 depicts the epitopes predicted using AAP, BCPred, and IEDB tools further filtered based on allergenicity and antigenicity properties. Antigenic regions common to both BCPred and AAP were subjected to IEDB Emini surface accessibility tool to predict peptides which were surface exposed. Predicted peptides were checked for flexibility and hydrophilicity using IEDB tools, Karplus and Schulz flexibility prediction and Parker hydrophilicity prediction. This yielded a total of 25 peptides as B-cell epitopes. To test such peptides for their potential as B-cell epitopes, they were checked for their allergenicity and antigenicity which yielded four epitopes with allergenicity score ≤0.06 and antigenicity score >1 as shown in Table 4.

Discussion

With the advancement in sequencing technologies, there has been remarkable progress in the vaccinology area, enabling researchers to finally move beyond the traditional vaccinology approach. With computational approaches, it is now feasible to access the entire antigenic repertoire of an organism. Reverse vaccinology (RV) approach to vaccine identification came into existence with addressing the problem of vaccine identification against Meningococcus B (Men B). Men B is a pathogen which was intractable to vaccine development using conventional vaccinology approach as its capsular polysaccharide is identical to a human self-antigen (Giuliani et al. 2006). Hitherto, RV has been practically applied against many pathogens (Maione et al. 2005, Thorpe et al. 2007). In the post-genomic era, power of omics data has been complemented by bioinformatics approaches which may lead to the discovery of unique antigens that may eventually improve existing vaccines. Many researchers have already proposed an epitope-based vaccine candidate against *C. jejuni* with their studies aimed at identifying vaccine candidates from specific proteins like cytolethal distending toxin (CDT), autotransporter protein CapA, polysaccharide capsules, etc. (Ingale and Goto 2014, Ashgar et al. 2007, Guerry et al. 2012). Developing killed WC vaccines is complicated by dearth of information on pathogenesis of *C. jejuni*, and development of flagellin subunit-based vaccines is complicated owing to antigenic diversity of *Campylobacter* flagellins. Perceiving the gaps in current efforts for vaccine development against *Campylobacter*, we have undertaken current study of genome wide screening of *C. jejuni* using an in silico approach, aimed at identifying potential vaccine candidates against this organism and expedite the efforts in this direction.

Currently, most vaccines are based on B-cell providing antibody-mediated immunity. However, T-cells confer long-lasting immunity while antibody-mediated immunity can be easily overcome by surge of antigens (Bacchetta et al. 2005). Cytotoxic CD8[+] T lymphocytes (CTL) hamper infectious agents from spreading by invading infected cells. Thus, in this study, we have proposed both B- and T-cell epitopes which could be experimentally tested for their efficacy in triggering

Table 2 Predicted potential T-cell epitopes, along with their interacting MHC-I and MHC-II alleles with an affinity of <500 nM and corresponding IC50 values (in parentheses)

| Epitope	Total no. of MHC peptide binders	MHC-I alleles	MHC-II alleles	Conservancy (%)
NTDQAQGTV	24	HLA-A*02:02 (495.45), HLA-A*02:03 (267.30), HLA-A*02:06 (82.22), HLA-A*02:11 (276.27), HLA-A*02:50 (32.01), HLA-A*03:01 (191.87), HLA-A*11:01 (20.61), HLA-A*31:01 (349.14), HLA-A*32:07 (15.78), HLA-A*32:15 (297.34), HLA-A*68:01 (239.88), HLA-A*68:02 (86.10), HLA-A*68:23 (6.57), HLA-A*69:01 (124.99), HLA-B*40:13 (253.47), HLA-C*05:01 (161.41), HLA-C*06:02 (197.61), HLA-C*07:01 (61.12), HLA-C*08:02 (87.33), HLA-C*12:03 (4.22), HLA-C*14:02 (145.08), HLA-C*15:02 (69.09)	HLA-DRB1*01:01 (138.04), HLA-DRB1*04:01 (328.85)	75
YIQDNFNFY	29	HLA-A*01:01 (3.94), HLA-A*02:02 (85.70), HLA-A*02:03 (126.47), HLA-A*02:06 (16.98), HLA-A*02:17 (190.84), HLA-A*03:01 (69.18), HLA-A*11:01 (69.82), HLA-A*25:01 (296.54), HLA-A*26:01 (306.63), HLA-A*26:02 (478.98), HLA-A*29:02 (49.29), HLA-A*30:02 (228.80), HLA-A*32:07 (41.50), HLA-A*32:15 (95.55), HLA-A*68:01 (232.27), HLA-A*68:23 (27.46), HLA-A*80:01 (87.30), HLA-B*15:01 (123.59), HLA-B*15:02 (94.02), HLA-B*15:03 (413.13), HLA-B*27:20 (134.12), HLA-B*35:01 (184.75), HLA-B*40:13 (138.33), HLA-C*03:03 (25.62), HLA-C*12:03 (13.59), HLA-C*14:02 (124.34)	HLA-DRB1*01:01 (0.60), HLA-DRB1*04:01 (158.49), HLA-DRB1*07:01 (144.21)	50
RSDEAQTNY	21	HLA-A*01:01 (108.39), HLA-A*11:01 (148.94), HLA-A*30:02 (245.73), HLA-A*31:01 (431.52), HLA-A*32:07 (11.91), HLA-A*32:15 (196.45), HLA-A*68:01 (367.28), HLA-A*68:23 (17.81), HLA-B*15:02 (117.06), HLA-B*15:02 (401.08), HLA-B*15:03 (153.85), HLA-B*15:17 (60.63),	HLA-DRB1*01:01 (8.07)	10.34

Table 2 (continued)

Epitope	Total no. of MHC peptide binders	MHC-I alleles	MHC-II alleles	Conservancy (%)
KSDEEMEKY	19	HLA-B*27:20 (6.48), HLA-B*40:13 (40.64), HLA-B*58:01 (415.24), HLA-C*05:01 (161.41), HLA-C*06:02 (275.30), HLA-C*07:01 (121.39), HLA-C*08:02 (388.29), HLA-C*12:03 (14.81) HLA-A*01:01 (98.40), HLA-A*02:02 (119.40), HLA-A*02:03 (98.17), HLA-A*02:06 (232.27), HLA-A*11:01 (144.54), HLA-A*30:02 (328.44), HLA-A*32:07 (18.80), HLA-A*32:15 (162.65), HLA-A*68:23 (25.99), HLA-B*15:17 (222.69), HLA-B*27:20 (5.57), HLA-B*40:13 (25.88), HLA-B*58:01 (363.33), HLA-C*05:01 (86.83), HLA-C*07:01 (491.10), HLA-C*12:03 (4.95), HLA-C*14:02 (315.22)	HLA-DRB1*01:01 (34.20), HLA-DRB1*07:01 (259.42)	100

Table 3 Population coverage of predicted epitopes based on MHC-I and MHC-II restriction data for epitopes NTDQAQGTV and YIQDNFNFY maximum population coverage by Europe

Population/area	Class I and II coverage (%)	
	NTDQAQGTV	YIQDNFNFY
World	81.07	85.27
East Asia	68.22	80.61
Northeast Asia	74.14	82.80
South Asia	84.25	85.70
Southeast Asia	63.61	70.72
Southwest Asia	72.88	70.55
Europe	85.99	90.81
East Africa	70.98	66.55
West Africa	69.64	75.91
Central Africa	75.76	69.61
North Africa	80.44	77.05
South Africa	85.53	77.44
West Indies	46.39	77.88
North America	78.72	84.08
Central America	1.34	15.10
South America	69.25	59.04
Oceania	63.73	64.30
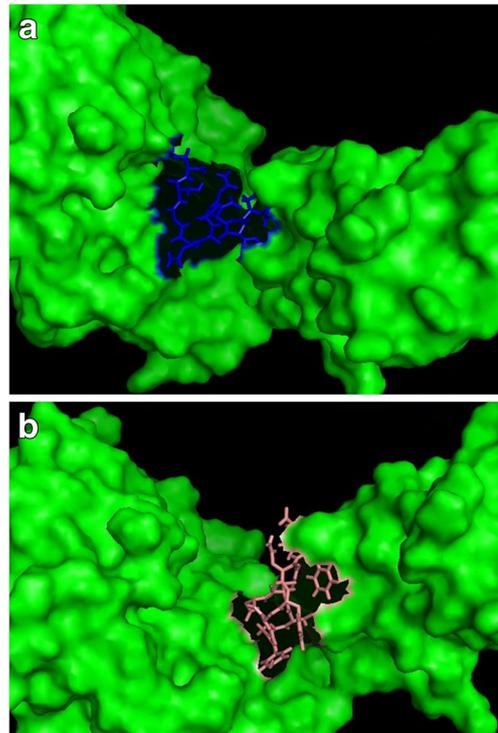
Fig. 2 Docked complexes of HLA-A*11:01 against predicted epitopes generated by Hex docking program. **a** Epitope NTDQAQGTV. **b** Epitope YIQDNFNFY
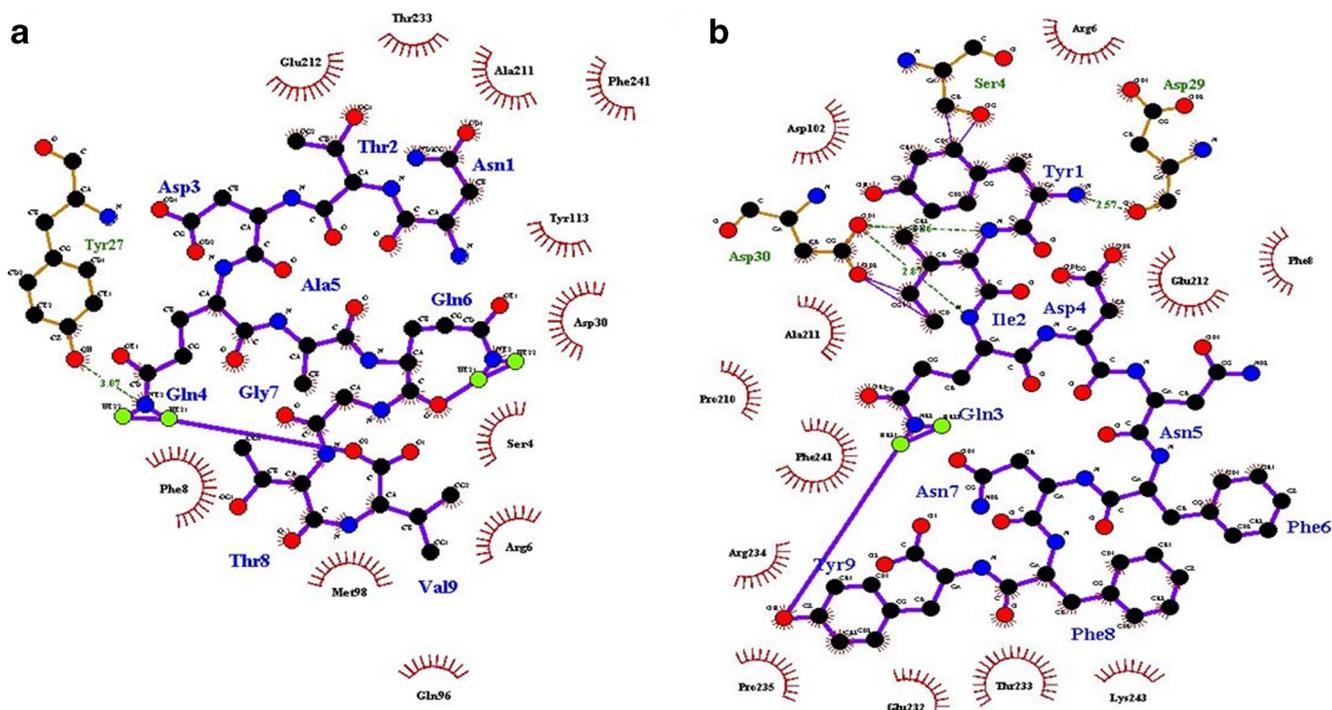
Fig. 3 Interactions involved in HLA-A*11:01 binding. **a** Epitope NTDQAQGTV. **b** Epitope YIQDNFNFY

humoral and cell-mediated immune responses. As described in schematic workflow diagram (Fig. 1), we have framed a set of criteria for identifying potential vaccine candidates which involves antigenicity, T-cell/B-cell processivity, interaction with HLA alleles, allergenicity, conservancy, and population coverage. Protective epitopes are not clearly defined for *C. jejuni*. Thus, while screening proteomic data, it is of utmost importance to select the proteins which can confer protection. To select such segments from the proteins, it is encouraged to select genomic segments with antigenic properties. Thus, antigenicity filter was employed at several stages of vaccine candidate identification task. Initially, the proteins with antigenicity score above threshold 0.7 were selected as antigenic. Identified B- and T-cell epitopes were also filtered on antigenicity criterion.

Physiochemical properties like flexibility, hydrophilicity, and solvent accessibility are distinctive features of B-cell epitopes. These features have been exploited in many B-cell epitope prediction programs (Li et al. 2014). Initially, based

on surface accessibility, flexibility and hydrophilicity criteria B-cell epitopes which could be proficiently processed by B lymphocytes were identified. NetCTL server predicted T-cell epitopes based on combined predictions of MHC class I binding, proteasomal C terminal cleavage, and TAP transport efficiency. *C. jejuni* strains are highly diverse which further complicates the vaccine development against this pathogen. Consequently, conservation of the epitopes at sequence level reveals that these regions are imperative from evolutionary point. Population coverage plays an essential role in vaccine development process. Our predicted peptides showed good population coverage in spite of the fact that in case of MHC-II data was only available for the alleles HLA-DRB1*01:01, HLA-DRB1*04:01, and HLA-DRB1*07:01. Though, all the predicted nonamers were interacting with the most common HLA allele HLA-DRB1*01:01 as shown in Table 2. For the predicted epitopes, in developing the world's highest population coverage was in Asian and African countries where the diarrheal incidence rate is

Table 4 Four most potential B-cell epitopes by combined predictions of AAP, BCPred, and IEDB tools (Emini Surface Accessibility, Karplus and Schulz flexibility, and Parker hydrophilicity) filtered based on their AllerHunter and VaxiJen score

Uniprot_Id	Protein name	Gene name	Start	End	B-cell epitope	Length	AllerHunter score	VaxiJen score
Q0PBW1	Putative TonB-dependent outer membrane receptor	Cj0178	648	660	YTGKAKRVNPNT	12	0.06	1.6624
Q0PAY6	Putative periplasmic protein	Cj0530	583	593	IYRKHSNSSNS	11	0.06	1.6432
Q0P9M1	Outer membrane component of efflux system (multidrug efflux system cmeDEF)	cmeD	220	228	NPQQEKSQN	9	0.05	1.2950
			327	336	RFSERKNKEE	10	0.06	1.3154

reported to be the highest, and in industrialized nations, it was highest for Europe and North America, aligning with the fact that maximum number of travelers to Asia and Africa are observed from these countries (Harris et al. 2011).

Further investigation in the data shows that epitope NTDQAQGTV has high antigenicity value (Table 2), but epitope YIQDNFNFY has a maximum of 29 MHC-interacting alleles. Epitope YIQDNFNFY was identified from cmeD which encodes for outer membrane component of multidrug efflux system cmeDEF. In a study, cmeC which is an essential outer membrane component of cmeABC multidrug efflux pump was proposed as a promising subunit vaccine candidate against *C. jejuni* infection using a chicken model (Zeng et al. 2010). cmeDEF also plays important role in antibiotic resistance against several antibiotics and toxic compounds. cmeABC and cmeDEF act synergistically in retaining cell viability and conferring antibiotic resistance (Akiba et al. 2006). Epitope sequence NTDQAQGTV has lower IC50 value of 20.61 nM for MHC supertype (HLA-A∗11:01) as compared to YIQDNFNFY which has an IC50 of 69.82 nM with HLA-A∗11:01. The results of computational docking studies coincide with the binding affinity values. NTDQAQGTV has a stronger affinity for HLA-A∗11:01 with binding energy of −386.53, while YIQDNFNFY binds in the groove of HLA-A∗11:01 with total energy −350.09. Epitope sequence YIQDNFNFY is more conserved at 100 % identity and has a high score as a processed peptide as evidenced from NetCTL score (Table 1). As seen in Table 3, population coverage analysis reveals that epitope YIQDNFNFY covers a large proportion of human population. Lowest coverage for this epitopes is in Central America (15.10 %) which is much higher when compared to population coverage of NTDQAQGTV in the same region being 1.34 %.

Four potential B-cell epitopes were predicted with their VaxiJen score (antigenicity) >1 and with AllerHunter score (allergenicity) ≤0.06 threshold. The epitopic sequence YTGKAKRVNPNT has the highest antigenicity of 1.6624 followed by IYRKHSNSSNS and RFSERKNKEE with antigenicity scores 1.6432 and 1.3154, respectively. Based on AllerHunter results, epitope sequence NPQQEKSQN has the highest possibility of being a non-allergen as marked by the lowest score 0.05, while other B-cell epitopes YTGKAKRVNPNT, IYRKHSNSSNS, and RFSERKNKEE have AllerHunter score of 0.06. Elicitation of effective immune responses depends on the specificity and diversity of the T-cell epitopes binding to HLA alleles. Due to highly polymorphic nature of MHC, it is desirable to identify peptides which can bind to many MHC alleles (Germain 1994). Our predicted T-cell epitopes NTDQAQGTV and YIQDNFNFY bind to more than 20 MHC alleles and have broad human population coverage.

HLA-A∗11:01 was selected for docking studies. The predicted T-cell epitopes interact with HLA-A∗11:01 with

varying affinities. From computational docking results, it was interpreted that the epitopes bind efficiently to the HLA-A∗11:01. It is believed that such a systematic computational pipeline for prediction of vaccine candidates when employed to *C. jejuni* proteome reveals epitopes that would be able to elicit an efficacious immune response.

There has been growing body of evidence which state the indispensable role of bioinformatics approaches in translational medicine (Ashgar et al. 2007; Guerry et al. 2012; Ingale and Goto 2014; Binnewies et al. 2006; Huang et al. 2002). As shown in Fig. 4, there was a sturdy decline in the protein search space at each step. It was noticed that there was significant reduction in the proteome size to be searched for vaccine identification. We started with the proteome size of 1623 proteins. Applying different filtration criteria at each step, we were left with eight antigenic sequences which were eventually tested for presence of vaccine candidates. In summary, omics-guided approaches and bioinformatics analyses offer broad potential for further developments in global health relevant novel therapeutics.

Conclusion

Traditional molecular immunology techniques for vaccine identification are time and labor consuming. A wide array of omics techniques, whole genome sequencing data, and novel bioinformatics approaches have substantially improved our systemic understanding of complex diseases. These techniques hold a greater potential to be utilized for rapid and reliable genome wide screening for identification of vaccine candidates, thus have hastened the pace of vaccine development to a great extent by significantly reducing the number of experimentally testable epitopes. Our predicted epitopes are prospective vaccine candidates on grounds of higher population coverage and interactions with many HLA alleles. In conclusion, an immunoinformatics-based approach was
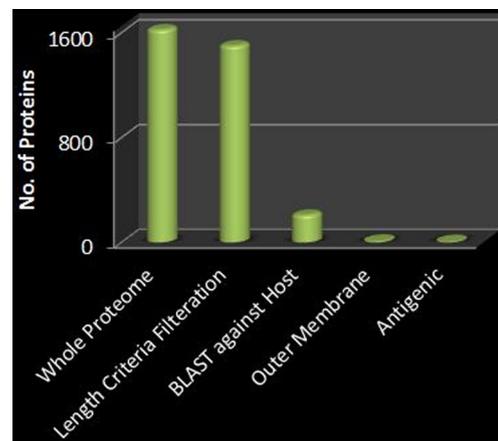
Fig. 4 Step-wise reduction in the total no. of proteins in search for the identification of vaccine candidates against *C. jejuni*

utilized for detection of protective antigens in *C. jejuni*, which may serve as potential vaccine candidates to control campylobacteriosis once validated experimentally in vitro and in vivo. This immunoinformatics-based approach can be applied to other hosts or other enteric pathogens.

Compliance with ethical standards

Conflict of interest We confirm that there are no conflicts of interest associated with this publication. Ethical Approval and Informed Consent statements are not applicable to our manuscript.

References

Akiba M, Lin J, Barton Y-W, Zhang Q (2006) Interaction of CmeABC and CmeDEF in conferring antimicrobial resistance and maintaining cell viability in Campylobacter jejuni. J Antimicrob Chemother 57:52–60

Albert MJ (2014) Vaccines against Campylobacter jejuni. Austin J Clin Immunol 1:1013

Alfredson DA, Korolik V (2007) Antibiotic resistance and resistance mechanisms in Campylobacter jejuni and Campylobacter coli. FEMS Microbiol Lett 277:123–132

Ashgar SSA, Oldfield NJ, Wooldridge KG, Jones MA, Irving GJ, Turner DPJ, Ala'aldeen DAA (2007) CapA, an autotransporter protein of Campylobacter jejuni, mediates sssociation with human epithelial cells and colonization of the chicken gut. J Bacteriol 189:1856–1865

Bacchetta R, Gregori S, Roncarolo M-G (2005) CD4+ regulatory T cells: mechanisms of induction and effector function. Autoimmun Rev 4: 491–496

Binnewies TT, Motro Y, Hallin PF, Lund O, Dunn D, La T, Hampson DJ, Bellgard M, Wassenaar TM, Ussery DW (2006) Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. Funct Integr Genomics 6:165–185

Brocchieri L, Karlin S (2005) Protein length in eukaryotic and prokaryotic proteomes. Nucleic Acids Res 33:3390–3400

Bui H-H, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A (2006) Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics 7:153–153

Bui H-H, Sidney J, Li W, Fusseder N, Sette A (2007) Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinformatics 8:361–361

Butt AM, Nasrullah I, Tahir S, Tong Y (2012) Comparative genomics analysis of mycobacterium ulcerans for the identification of putative essential genes and therapeutic candidates. PLoS One 7:e43080

Chen J, Liu H, Yang J, Chou KC (2007) Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. Amino Acids 33:423–428

Dekeyser P, Gossuin-Detrain M, Butzler JP, Sternon J (1972) Acute enteritis due to related vibrio: first positive stool cultures. J Infect Dis 125:390–392

Doro F, Liberatori S, Rodrãguez-Ortega MJ, Rinaudo CD, Rosini R, Mora M, Scarselli M, Altindis E, D'aurizio R, Stella M, Margarit I, Maione D, Telford JL, Norais N, Grandi G (2009) Surfome analysis as a fast track to vaccine discovery: identification of a novel protective antigen for group b streptococcus hypervirulent strain COH1. Mol Cell Proteomics 8:1728–1737

Doytchinova I, Flower D (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 8:4

El-Manzalawy Y, Dobbs D, Honavar V (2008) Predicting linear B-cell epitopes using string kernels. J Mol Recognit 21:243–255

Emini EA, Hughes JV, Perlow DS, Boger J (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. J Virol 55:836–839

Friedman CR, Neimann J, Wegener HC, Tauxe RV (2000) Epidemiology of Campylobacter jejuni infections in the United States and other industrialized nations. In: Campylobacter. ASM International, Washington, pp 121–138

Germain RN (1994) MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation. Cell 76:287–299

Giuliani MM, Adu-Bobie J, Comanducci M, Aricã B, Savino S, Santini L, Brunelli B, Bambini S, Biolchi A, Capecchi B, Cartocci E, Ciucchi L, Di Marcello F, Ferlicca F, Galli B, Luzzi E, Masignani V, Serruto D, Veggi D, Contorni M, Morandi M, Bartalesi A, Cinotti V, Mannucci D, Titta F, Ovidi E, Welsch JA, Granoff D, Rappuoli R, Pizza M (2006) A universal vaccine for serogroup B meningococcus. Proc Natl Acad Sci U S A 103:10834–10839

Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR (2011) Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. Nucleic Acids Res 39:D913–D919

Guan P, Doytchinova IA, Zygouri C, Flower DR (2003) MHCPred: a server for quantitative prediction of peptide-MHC binding. Nucleic Acids Res 31:3621–3624

Guerry P, Poly F, Riddle M, Maue AC, Chen Y-H, Monteiro MA (2012) Campylobacter polysaccharide capsules: virulence and vaccines. Front Cell Infect Microbiol 2:7

Harris JA, Roy K, Woo-Rasberry V, Hamilton DJ, Kansal R, Qadri F, Fleckenstein JM (2011) Directed evaluation of enterotoxigenic Escherichia coli autotransporter proteins as putative vaccine candidates. PLoS Negl Trop Dis 5:e1428

Huang S-H, Triche T, Jong AY (2002) Infectomics: genomics and proteomics of microbial infections. Funct Integr Genomics 1:331–344

Ingale A, Goto S (2014) Prediction of CTL epitope, in silico modeling and functional analysis of cytolethal distending toxin (CDT) protein of Campylobacter jejuni. BMC Res Notes 7:92

Karplus PA, Schulz GE (1985) Prediction of chain flexibility in proteins. Naturwissenschaften 72:212–213

Ketley JM (1997) Pathogenesis of enteric infection by campylobacter. Microbiology 143:5–21

Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics 8:424–424

Laskowski RA, Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model 51: 2778–2786

Li X, Yang H-W, Chen H, Wu J, Liu Y, Wei J-F (2014) In silico prediction of T and B cell epitopes of Der f 25 in Dermatophagoides farinae. Int J Genomics 2014:10

Macindoe G, Mavridis L, Venkatraman V, Devignes M-D, Ritchie DW (2010) HexServer: an FFT-based protein docking server powered by graphics processors. Nucleic Acids Res 38:W445–W449

Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, Scarselli M, Tettelin H, Brettoni C, Iacobini ET, Rosini R, D'Agostino N, Miorin L, Buccato S, Mariani M, Galli G, Nogarotto R, Dei VN, Vegni F, Fraser C, Mancuso G, Teti G, Madoff LC, Paoletti LC, Rappuoli R, Kasper DL, Telford JL, Grandi G (2005) Identification of a universal group B streptococcus vaccine by multiple genome screen. Science (New York, NY) 309:148–150

Moriel DG, Scarselli M, Serino L, Mora M, Rappuoli R, Masignani V (2008) Genome-based vaccine development: a short cut for the future. Hum Vaccin 4:184–188

Muh HC, Tong JC, Tammi MT (2009) AllerHunter: a SVM-pairwise system for assessment of allergenicity and allergic cross-reactivity in proteins. PLoS One 4:e5861

O'ryan M, Vidal R, Del Canto F, Carlos SJ, Montero D (2015) Vaccines for viral and bacterial pathogens causing acute gastroenteritis: part II: vaccines for shigella, salmonella, enterotoxigenic E. coli (ETEC) enterohemorragic E. coli (EHEC) and Campylobacter jejuni. Hum Vaccin Immunother 11:601–619

Parker JMR, Guo D, Hodges RS (1986) New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and x-ray-derived accessible sites. Biochemistry 25:5425–5432

Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, Basham D, Chillingworth T, Davies RM, Feltwell T, Holroyd S, Jagels K, Karlyshev AV, Moule S, Pallen MJ, Penn CW, Quail MA, Rajandream MA, Rutherford KM, Van Vliet AHM, Whitehead S, Barrell BG (2000) The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences. Nature 403:665–668

Peters B, Sette A (2005) Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. BMC Bioinformatics 6:132–132

Rinaudo CD, Telford JL, Rappuoli R, Seib KL (2009) Vaccinology in the genome era. J Clin Invest 119:2515–2525

Schrodinger, Llc (2010) The PyMOL molecular graphics system, Version 1.3r1

Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, Kloetzel PM, Rammensee HG, Schild H, Holzhütter HG (2005) Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. Cell Mol Life Sci 62:1025–1037

Thevenet P, Shen Y, Maupetit J, Guyon FDR, Derreumaux P, TuffeRy P (2012) PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. Nucleic Acids Res 40(TheVenet P):W288–W293

Thakur S, Zhao S, Mcdermott PF, Harbottle H, Abbott J, English L, Gebreyes WA, White DG (2010) Antimicrobial resistance, virulence, and genotypic profile comparison of Campylobacter jejuni and Campylobacter coli isolated from humans and retail meats. Foodborne Pathog Dis 7:835–844

Thorpe C, Edwards L, Snelgrove R, Finco O, Rae A, Grandi G, Guilio R, Hussell T (2007) Discovery of a vaccine antigen that protects mice from Chlamydia pneumoniae infection. Vaccine 25:2252–2260

Yu NY, Wagner JR, Laird MR, Melli G, Rey SB, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FSL (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615

Zeng X, Xu F, Lin J (2010) Development and evaluation of CmeC subunit vaccine against Campylobacter jejuni. J Vaccines Vaccination 1:112