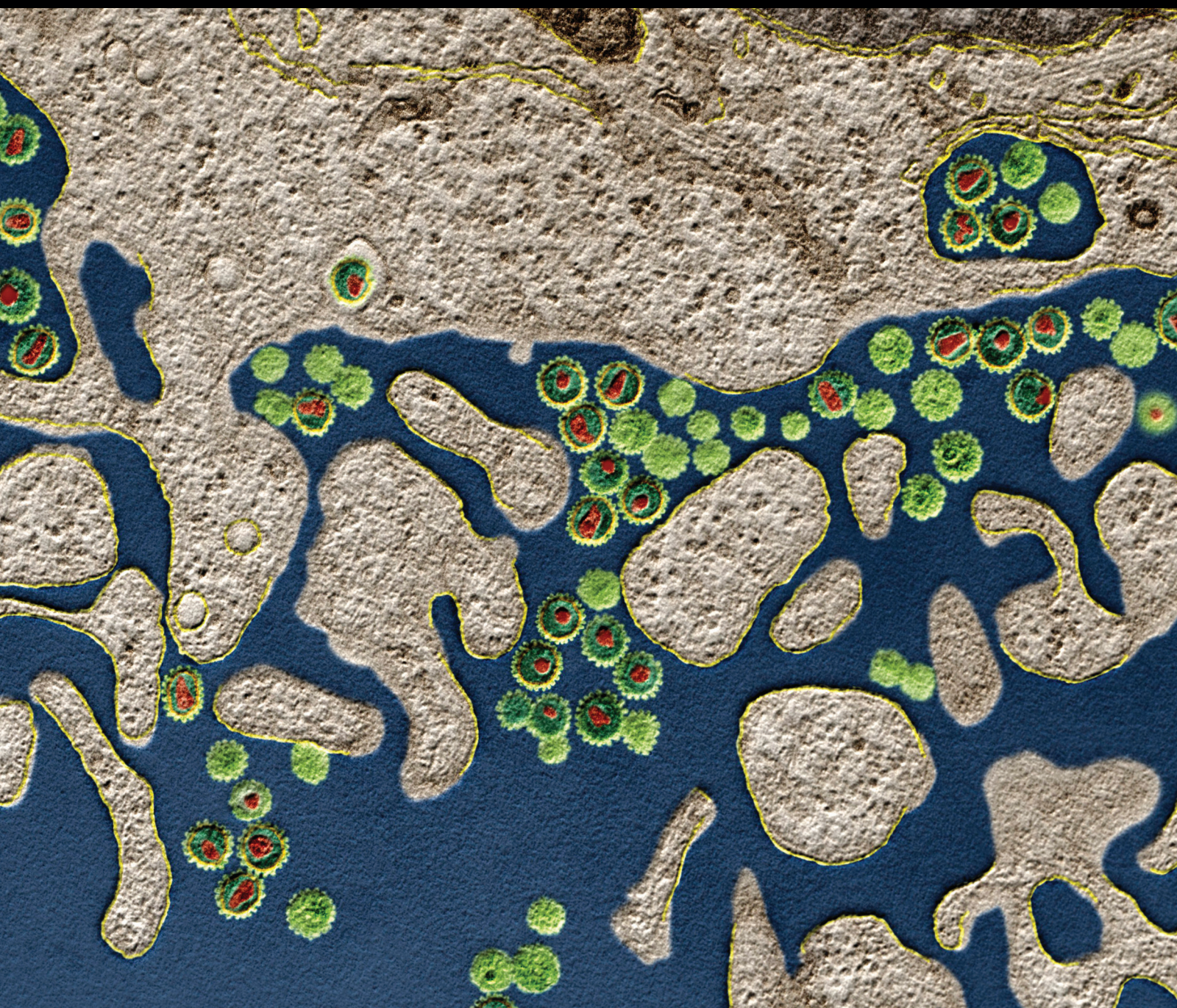
Journal of Immunology Research

Peptide-Based Immunotherapeutics and Vaccines 2017

Lead Guest Editor: Pedro A. Reche
Guest Editors: Darren R. Flower, Masha Fridkis-Hareli, and Yoshihiko Hoshino



Peptide-Based Immunotherapeutics and Vaccines 2017

Journal of Immunology Research

Peptide-Based Immunotherapeutics and Vaccines 2017

Lead Guest Editor: Pedro A. Reche
Guest Editors: Darren R. Flower, Masha Fridkis-Hareli, and Yoshihiko Hoshino

Editorial Board

Contents

Hindawi
Journal of Immunology Research
Volume 2018, Article ID 4568239, 2 pages
https://doi.org/10.1155/2018/4568239

Editorial

Peptide-Based Immunotherapeutics and Vaccines 2017

Pedro Reche,[1] **Darren R. Flower**,[2] **Masha Fridkis-Hareli**,[3] **and Yoshihiko Hoshino**[4]

[1]*Department of Immunology & O2, Facultad de Medicina, Universidad Complutense, 28040 Madrid, Spain*
[2]*School of Life and Health Sciences, Aston University, Birmingham B4 7ET, UK*
[3]*ATR, LLC, Worcester, MA 01606, USA*
[4]*Department of Mycobacteriology, National Institute of Infectious Diseases, Higashi-Murayama, Tokyo 189-0002, Japan*

Correspondence should be addressed to Pedro Reche; parecheg@med.ucm.es

Received 4 February 2018; Accepted 4 February 2018; Published 15 July 2018

The increasing understanding of the immune system and the critical role of antigenic epitopes in eliciting robust immune responses has led to the development of peptide vaccines. Peptide-based vaccines or—more technical—epitope ensemble vaccines represent an alternative approach to the discovery of disease-specific prophylactic and therapeutic vaccines, distinct from other vaccine moieties, such as attenuated or killed whole pathogen vaccines, subunit or toxoid vaccines, and carbohydrate-based vaccines. Epitopes represent the relevant part of the antigen recognized by T and/or B cells, mediating adaptive immunity. Consequently, the major potential of epitope vaccine ensembles is that of inducing desirable T and B cell-mediated immune responses. The risk of causing pathogenic or off-target responses with epitope vaccines is thus much lower than with convectional vaccines and can, on that basis, be regarded as safer. Epitope ensemble vaccines are also very versatile and can be formulated as synthetic peptides or encoded as DNA and RNA formulations. Currently, there are many epitope vaccines under development, spanning a wide range of diseases, including chronic viral infections and therapeutic anticancer vaccines, but as yet none are available. The case for epitope ensemble vaccines is nonetheless compelling, and the research community remains certain it is only time that separates us from a viable and deployable vaccine.

Developing a successful peptide-based vaccine involves identifying disease-specific epitopes inducing protective immunity and tackling a number of steps, such as determining appropriate means of epitope delivery and overcoming the intrinsically low immunogenicity of isolated epitopes. Following the success of previous special issues on peptide-based vaccines and immunotherapeutics, we present a new edition of the special issue in which we have incorporated seven original articles and two reviews, addressing various aspects of peptide-based vaccine design.

The reviews in the issue deal with epitope databases and epitope prediction. Peptide databases with information on T and B cell epitopes and peptide binding to MHC molecules are important tools for analyzing immune responses, benchmarking predictive methods, generating new ones, and developing peptide-based immunotherapeutics [1]. Currently, there are a number of online resources providing this type of information [2] but the Immune Epitope Database (IEDB) is the largest and most comprehensive epitope database. Thus, in this issue, we are pleased to feature a review by W. Fleri et al., in which the authors describe how the data are entered and retrieved from IEDB. B and T cell epitopes can be predicted from the relevant antigens with the help of bioinformatics tools. In this issue, J. L. Sanchez-Trincado et al. analyzed aspects of antigen recognition by T and B cells that are relevant for epitope prediction and provided a systematic and inclusive review of available tools, paying particular attention to their foundations. In the review, the authors also provide arguments on why B cell epitope prediction is less accurate and practical than T cell prediction and introduce solutions to solve some of the problems associated with epitope-vaccine design.

The original articles in this special issue include two papers focusing on computational vaccine design. In one, A. R. Oany et al. proposed a peptide vaccine candidate for

shigellosis consisting of four predicted cytotoxic T cell (CTL) epitopes from the SigA antigen, which is known to be highly immunogenic. These four candidates are conserved among *Shigella* species and provide wide population coverage. In addition, the authors identified that MHC II molecules could also present two of the CTL epitopes and thus stimulate T helper responses. In the other, J. Alonso-Padilla et al. have expanded on an emerging strategy that relies on the use of experimentally defined T-cell epitopes [3] and formulated a prophylactic epitope vaccine against EBV infection that includes both T and B cell epitopes. The T cell component consists of experimentally defined CD8 and CD4 T cell epitopes from various EBV antigens that are conserved and can be presented by a large number of human MHC molecules, while the B cell component includes experimental B cell epitopes mapping on the ectodomain of EBV envelope proteins and exhibiting a high degree of flexibility and solvent accessibility.

The experimental articles in the issue range widely. We have a work by M. Niki et al. which aims to identify appropriate antigens for tuberculosis (TB) vaccines. The study is an extension of a previous work in which the authors did a cross-sectional assay in TB patients [4]. Here, the authors used a different cohort and did a longitudinal assay, finding immunoglobulin responses to antigens that correlated to several clinical parameters. These results provide insights into the development of a novel TB vaccine inducing protective humoral immunity.

Dendritic cell- (DC) and peptide-based immunotherapies often go hand in hand. In this issue, J. Lo et al. investigate on whether DCs unpulsed or pulsed with antigenic dominant determinants (DD), subdominant determinants (SD), and ignored determinants (ID) could prevent type 1 diabetes (T1D) in a mouse model. The authors found that diabetes was significantly delayed by DCs pulsed with SD or ID peptides. Moreover, they also found that Tregs from DC-treated mice proliferated more actively and showed enhanced immunosuppressive activities. Overall, this study demonstrates that DC therapy leads to long-lasting immunomodulatory effects in an antigen-dependent manner, providing support for DC-guided peptide-based interventions for autoimmune diabetes.

Toxicity is a handicap for vaccine design and, here, A. Latanova et al. present a work addressing this issue. In the study, the authors fused a flaviviral leader peptide to reverse transcriptase (RT), a crucial target of immunotherapy against drug-resistant HIV-1. This fusion allowed RT secretion and reduced its toxicity and ability to induce oxidative stress, with no major effects on its immunogenicity. Subsequently, the authors proposed the use of leader peptides to increase safety of RT-based DNA vaccines. Genetic diversity of pathogens remain also a substantial obstacle for vaccine design. V. S. Kichatova et al. addressed this subject for human hepatitis C virus (HCV). These authors characterized the occurrence of IFN resistance-conferring mutations in HCV isolates circulating in the Russian Federation, identifying that the spread of viral variants was linked to mutations on HCV-specific CTL epitopes in association with the immunogenetic background of HCV-infected individuals. These results are useful in identifying those individuals in need of IFN-free treatments and for developing epitope-based vaccines that circumvent viral immune escape.

Finally, we include a work by I. Soria et al. showing that combining viral-specific B and T cell epitopes onto appropriated structures can increase immunogenicity and enhance protection. These authors work with foot-and-mouth disease virus (FMDV), which has a high morbidity in cloven-hoofed animals, like cattle and swine. In cattle, there are known FMDV-specific B and T cell epitopes that could be used for safer and more effective vaccines. However, immunization with linear synthetic peptides encompassing the epitopes has failed to induce protection in cattle. In contrast, in the study, the authors immunized cattle with a dendrimeric peptide structures consisting of 4 copies of a peptide encompassing a B cell epitope linked through thioether bonds to a single copy of a CD4 T cell epitope. As a result, I. Soria et al. obtained that these dendrimeric peptides elicited humoral and cellular immune responses that conferred partial protection against heterologous virus challenge.

In conclusion, the articles included in this special issue examined relevant aspects of peptide-based vaccines, and we trust that readers shall find them both interesting and motivating.

Pedro Reche
Darren R. Flower
Masha Fridkis-Hareli
Yoshihiko Hoshino

References

[1] D. R. Flower, "Databases and data mining for computational vaccinology," *Current Opinion in Drug Discovery & Development*, vol. 6, no. 3, pp. 396–400, 2003.

[2] Y. He and Z. Xiang, "Databases and in silico tools for vaccine design," *Methods in Molecular Biology*, vol. 993, pp. 115–127, 2013.

[3] Q. M. Sheikh, D. Gatherer, P. A. Reche, and D. R. Flower, "Towards the knowledge-based design of universal influenza epitope ensemble vaccines," *Bioinformatics*, vol. 32, no. 21, pp. 3233–3239, 2016, Epub 2016 Jul 10.

[4] M. Niki, M. Suzukawa, S. Akashi et al., "Evaluation of humoral immunity to *Mycobacterium tuberculosis*-specific antigens for correlation with clinical status and effective vaccine development," *Journal of Immunology Research*, vol. 2015, Article ID 527395, 13 pages, 2015.

Hindawi
Journal of Immunology Research
Volume 2018, Article ID 4928757, 11 pages
https://doi.org/10.1155/2018/4928757

Research Article

Longitudinal Evaluation of Humoral Immunity and Bacterial and Clinical Parameters Reveals That Antigen-Specific Antibodies Suppress Inflammatory Responses in Active Tuberculosis Patients

Mamiko Niki,[1] Takashi Yoshiyama,[2] Yuji Miyamoto,[3] Masao Okumura,[2] Makoto Niki,[1] Ken-ichi Oinuma,[1] Yukihiro Kaneko,[1] Sohkichi Matsumoto,[4] Yuka Sasaki,[2] Hideo Ogata,[2] Hajime Goto,[2] Shoji Kudoh,[2] and Yoshihiko Hoshino[3]

[1]*Department of Bacteriology, Osaka City University Graduate School of Medicine, Abeno, Osaka 545-8585, Japan*
[2]*Division of Respiratory Medicine, Fukujuji Hospital, Japan Anti-Tuberculosis Association, Matsuyama, Kiyose, Tokyo 204-8522, Japan*
[3]*Department of Mycobacteriology, Leprosy Research Center, National Institute of Infectious Diseases, Aoba, Higashi-Murayama, Tokyo 189-0002, Japan*
[4]*Department of Bacteriology, Niigata University Graduate School of Medicine, Niigata 951-8510, Japan*

Correspondence should be addressed to Yoshihiko Hoshino; yhoshino@nih.go.jp

Received 19 May 2017; Revised 6 November 2017; Accepted 31 December 2017; Published 4 July 2018

Academic Editor: Senthamil R. Selvan

A novel tuberculosis vaccine to replace BCG has long been desired. However, recent vaccine trials focused on cell-mediated immunity have failed to produce promising results. It is worth noting that most commercially available successful vaccines rely on humoral immunity. To establish a basic understanding of humoral immunity against tuberculosis, we analyzed and evaluated longitudinal levels and avidity of immunoglobulin to various tuberculosis antigens compared with bacterial and clinical parameters during treatment. We found that levels of IgG antibodies against HrpA and HBHA prior to treatment exhibited a positive correlation with bacterial burden. Analysis of changes in CRP during treatment revealed an association with high levels of specific IgG and IgA antibodies against mycobacterial antigens. Levels of CRP prior to treatment were negatively associated with IgG avidity to CFP-10 and MDP1 and IgA avidity to HrpA, while IgA avidity to MDP1 and Acr exhibited a negative correlation with CRP levels after 60 days of treatment. These results may provide insight for the development of a novel tuberculosis (TB) vaccine candidate to induce protective humoral immunity against tuberculosis.

1. Introduction

Tuberculosis remains one of the most prevalent infectious diseases worldwide and is caused by *Mycobacterium tuberculosis* (Mtb). There were 1.4 million tuberculosis (TB) deaths and 10.4 million new TB cases in 2015. Approximately, one-third of the world's population is latently infected with Mtb, which represents a huge reservoir of future disease progression and transmission. Mtb is transmitted through the air from a person with active TB to a healthy individual. In addition, Mtb has no environmental or animal reservoirs and is believed to have coevolved with humans [1]. Therefore, an effective Mtb vaccine to prevent infection is the best strategy to eradicate tuberculosis.

The BCG vaccine has been the only available way to combat tuberculosis long before the development of antibiotics [2]. Although the BCG vaccine is effective for the prevention of children's disseminated tuberculosis, it has limited protective capacity on the development of adult pulmonary tuberculosis caused by the reactivation of persistent Mtb. With increasing numbers of cases involving HIV- and TB-coinfected people and multidrug-resistant

TB, the development of a more effective vaccination strategy is sorely needed.

As Mtb is an intracellular pathogen, the activation of cell-mediated immunity (CMI) characterized by IFN-γ-producing $CD4^+$ T cells is regarded as crucial for TB immunity [3]. Research on vaccines that induce CMI has dominated much of the conventional development effort, since antibody-mediated immunity has been considered to play little role in the outcome of Mtb infection [4]. To date, several candidate vaccines targeting the induction of CMI have reached clinical trials. One of these new vaccines, namely, MVA85A, is an attenuated vaccinia virus expressing Mtb antigen Ag85A, which induces high levels of antigen-specific $CD4^+$ and $CD8^+$ T cells in a murine model [5]. Although MVA85A achieved successful results in mouse models and phase I human clinical trials, the final outcome of the trial was found to exhibit no detectable improvement of protection against TB [6, 7]. Recently, a first-in-human phase I trial was conducted to investigate AERAS-422, a recombinant BCG overexpressing Mtb antigens and mutant perfringolysin derived from *Clostridium perfringens* [8]. However, high-dose AERAS-422 vaccination was found to elicit reactivation of the varicella-zoster virus (VZV), possibly due to negative regulation of immune control of latent VZV induced by the elevated IFN-γ production [9]. It is obvious that the conventional approach to targeting CMI against TB is insufficient.

In contrast to the research on CMI, many of the studies on humoral immune responses against Mtb antigens focused largely on their use in the diagnosis of TB, since some of the studies indicated that the serum antibody levels against Mtb antigens correlate with the degree of bacterial load [10, 11]. However, accumulating experimental evidence suggests that humoral immunity can modulate the immune response to intracellular pathogens [12–16]. In addition, studies on vaccines with protective efficacy based on antibody-mediated immunity against some of these pathogens have been reported [17–21]. Therefore, humoral immunity has been consistently highlighted as an important component of protective immune responses to Mtb [22]. As several reports have revealed a potential role of specific antibodies in host defense against Mtb [23–26], vaccination that induces Mtb-specific antibodies in the airway mucosa could be an effective strategy for protection against primary infection prior to Mtb entry into the lung.

In addition to the quantity of antibodies, the avidity of antibodies seems to be an important contributing factor to the protective capacity of vaccines. Antibody avidity is the functional affinity of multivalent antibody to bind multivalent antigens. It can be used to determine the net antigen binding force of a heterogeneous population of antibodies and has been used as a marker of B cell maturation during viral and bacterial infection [27–30]. In many infectious diseases including TB, elevated antibody avidity is observed in patients with chronic or severe conditions [31]. Although high-avidity antibodies are found to be crucial for the protection conferred by vaccines against many pathogens [32–35], there has been little investigation into the role of antibody avidity of anti-Mtb antibodies in protection against TB.

We previously demonstrated that serum levels of Mtb antigen-specific IgA, not IgG, correlated with clinical statuses of TB patients, suggesting that specific IgA antibodies could play a role in protection from disease [36]. In this study, we observed a relationship between clinical parameters related to TB severity and a change in antibody levels and antibody avidity to Mtb antigens, both before and after treatment.

2. Subjects and Methods

2.1. Participants. Patients of Fukujuji Hospital, Tokyo, Japan, were consecutively enrolled, after giving written informed consent, from April 2010 to March 2013. A total of 205 patients were recruited for this study. Patients were diagnosed as active-phase tuberculosis by clinical symptoms, chest X-ray images, and bacterial cultures. When blood samples both before and after treatment were available, they were included in the analysis. A total of thirty-three patients (age; 55.6 ± 16.4 yrs, male: 66.7%) were analyzed (Table 1). All patients took Japanese standard medications for tuberculosis (RFP + INH + EB(SM) + PZA for 2 months and RFP + INH for an additional 4 months) [37]. No patients dropped out of the treatment during the clinical course, and no patients were relapsed by the time of the analysis. There were no deceased patients during this analysis. The following information was obtained from all patients at the time of enrollment: history of prior TB disease, work history in any healthcare setting or recent exposure to a patient with active TB, and other TB risk factors, such as having immunodeficiency disorders or taking immunosuppressive drugs. We used the same inclusion/exclusion criteria as in a previous study [36]. Information on previous medical history and clinical signs and symptoms were also collected as previously described [36]. "Smear at entry" (entry = point of diagnosis before treatment) indicates the number of acid-fast bacilli inspected in the sputum smear taken at entry. The severity was subdivided as 0 (no acid-fast bacilli (AFB) on smear), ± (1-2 AFB per 300 fields), 1+ (1–9 AFB per 100 fields), 2+ (more than 10 AFB per 100 fields), and 3+ (more than 10 AFB per field). Several routine laboratory tests, including serum concentration of "C-reactive protein (CRP) at entry" and "CRP after 60 days" of treatment, were simultaneously performed. Blood sample collection was performed before treatment and after treatment for the analysis of immunoglobulin levels and avidity. The research protocol was approved by the Institutional Review Boards of Osaka City University Graduate School of Medicine, Osaka, Japan, and Fukujuji Hospital, Tokyo, Japan, and by the Research Ethics Committee of the National Institute of Infectious Disease, Tokyo, Japan.

2.2. Measurement of Serum Antibody Levels. Concentrations of IgG and IgA antibodies against Mtb were determined by ELISA using recombinant proteins as previously described with slight modification [36]. Ninety-six well microplates (Sumilon Type H, LMS, Tokyo, Japan) were coated with each recombinant antigen in bicarbonate buffer, pH 9.6 overnight at 4°C. The plates were blocked with phosphate buffered saline (PBS) containing 0.05% Tween 20 and 5% skim milk

Table 1: Individual patient characteristics. Severity of smear at entry was subdivided as 0 (no acid-fast bacilli (AFB) on smear), ± (1-2 AFB per 300 fields), 1+ (1–9 AFB per 100 fields), 2+ (more than 10 AFB per 100 fields), and 3+ (more than 10 AFB per fields).

ID	Age	Sex	Smear at entry	CRP at entry (mg/dl)	CRP after 60 days (mg/dl)
1	83	M	1	3.68	0
2	63	F	3	2.39	0.39
3	61	F	2	10.4	3.55
4	64	M	2	0.17	0
5	36	F	2	4.93	1.14
6	43	M	3	0.6	0.18
7	66	M	3	12.79	13.09
8	33	F	2	0.02	0.02
9	56	M	2	0.04	0.05
10	63	M	3	10.45	8.7
11	71	M	3	0.69	0
12	71	M	3	11.31	3.91
13	68	M	3	5.12	5.27
14	69	F	2	2.4	0.17
15	57	M	3	7.36	2.99
16	54	M	3	16.52	1.15
17	83	M	2	2.71	1.28
18	32	M	2	0.11	0.11
19	47	M	3	11.47	4.67
20	46	M	2	0.44	0.06
21	64	M	2	0.85	1.59
22	51	M	2	1.78	0.2
23	68	M	3	1.9	0.83
24	78	F	3	4.2	1.73
25	76	F	2	2.87	1.76
26	38	F	3	0.56	0.5
27	27	F	1	4.73	0.27
28	34	F	2	1.43	0.01
29	54	M	2	6.12	2.8
30	62	F	3	18.63	6.52
31	59	M	2	6.12	7.92
32	27	M	3	4.04	1.66
33	31	M	1	1.76	0.05

for 12 hr at 4°C and washed four times with PBS containing 0.05% Tween 20. Human serum samples diluted 1 : 200 in PBS containing 0.05% Tween 20 and 0.5% skim milk were then added in duplicate (IgG) or triplicate (IgA) to the antigen-coated wells and incubated for 12 hr at 4°C. After washing the wells, HRP-conjugated anti-human IgG or IgA antibodies were added at a 1 : 2000 or 1 : 1000 dilution, respectively. Following one-hour incubation at 37°C, the plates were washed four times before 100 μl of SureBlue reserve TMB was added to each well. The reactions were stopped after 10 min by adding 50 μl of 0.1 M HCl, and absorbance was measured at 450 nm using a Multiskan Spectrophotometer (Thermo Fisher Scientific, Yokohama, Japan). The results of the IgG-ELISA were expressed as absorbance at 450 nm, whereas results of the IgA-ELISA were expressed as ELISA-Index, $S/(B + 3SD)$, where S is the average OD value of the triplicate test samples and $B + 3SD$ [38] corresponds to the average OD value of the triplicate negative controls (B) plus three times the standard deviation (SD).

2.3. Avidity ELISA. Antibody avidity is a measure of the overall accumulated strength of the interaction between multiple antigenic epitopes and a multivalent antibody. For the measurement of the avidity of antibody, an incubation step with 7 M urea for 15 min after the serum incubation to elute the low-avidity antibodies was added to the ELISA assay procedure described above. Avidity indexes were obtained by calculating the ratio of the antibody levels measured by ELISA with and without urea treatment.

2.4. Reagents and Recombinant Protein Preparation. pET-21b, pET-22b, Luria-Bertani (LB) medium, and carbenicillin were from Sigma (St. Louis, MO, USA); isopropyl-1-thio-beta-D-galactopyranoside and Ni-NTA agarose were from Qiagen (Gaithersburg, MD, USA); skim milk was from Morinaga (Tokyo, Japan); horseradish peroxidase-conjugated anti-human IgG or IgA antibodies were from Dako (Carpinteria, CA, USA); SureBlue reserve TMB microwell peroxidase substrate was from KPL (Gaithersburg, MD, USA). Expression and purification procedures for recombinant mycobacterial antigens (ESAT-6, CFP-10, MDP1, Ag85A, Acr, HBHA, and HrpA [Acr2]) were described previously [36].

2.5. Statistical Analysis. Spearman's rank correlation coefficient was used to determine the correlation between ELISA values and the severity of clinical status values. All analyses were performed using online statistics calculators (http://www.socscistatistics.com/tests/Default.aspx, http://vassarstats.net/index.html, http://molpath.charite.de/cutoff/index.jsp). The threshold of significance was set at $p < 0.05$.

3. Results

3.1. Results of Clinical Parameters before and after Tuberculosis Treatment. We collected bacterial and clinical parameters as severity of smear at entry, CRP levels (mg/dl) at entry, and CRP levels (mg/dl) after 60 days of treatment (Table 1). We also measured levels of humoral immunity as IgG level, IgG avidity index, IgA level, and IgA avidity index (Supplemental Table 1).

3.2. Measurement and Comparison of Serum IgG Levels and IgG Avidity to Various Mtb Antigens before and after Treatment. We evaluated whether TB treatment affects serum antibody levels and antibody avidity indices for various Mtb antigens. A previous study showed that the levels of IgG against certain antigens decreased after the initiation of treatment [36], although it was also reported that the antibody response is heterogeneous and varies by individual, type of antigen, severity of the disease, and bacterial load [39, 40]. In our study, we observed a significant decrease in IgG levels against Acr and HrpA and in IgG avidity levels against

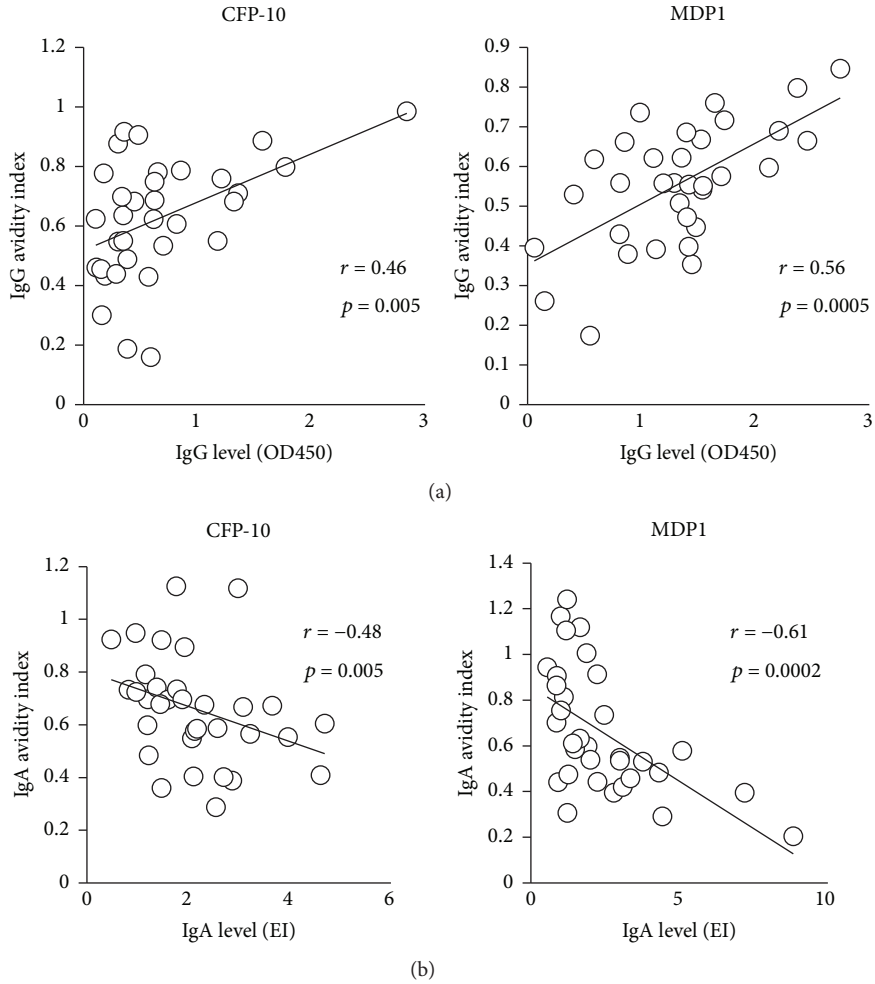


Figure 1: (a) Correlation between IgG levels and IgG avidity index against CFP-10 and MDP1 before treatment. (b) Correlation between IgA levels and IgA avidity index against CFP-10 and MDP1 before treatment.

MDP1 and Ag85A, during treatment (Supplemental Figure 1A–1B). On the other hand, neither IgA levels nor IgA avidity indices showed significant differences when compared before and after treatment (Supplemental Figure 2A–2B). Correlations between the change of serum antibody levels and antibody avidity levels during treatment were also analyzed. We found a statistically significant positive correlation between serum IgG levels and IgG avidity levels against CFP-10 and MDP1 before treatment (Figure 1(a)). Conversely, analysis of the relationship between serum IgA levels and IgA avidity against CFP-10 and MDP1 showed a negative correlation before treatment (Figure 1(b)). After treatment, only serum IgG level and avidity level against CFP-10 showed a positive correlation (Figure 2(a)), whereas almost all IgA tested was found to have a negative correlation between their antibody levels and the avidity levels (Figure 2(b)).

3.3. Analysis of the Relationship between Bacterial Load and Serum Mtb Antigen-Specific Antibody. To investigate whether the bacterial load affected the quantity or avidity of Mtb antigen-specific antibody, we compared the antibody levels and their avidity levels with "smear at entry" value. We found that HBHA- and HrpA-IgG levels before treatment showed a positive relationship with "smear at entry" (Figure 3(a)). We also observed that patients with high "smear at entry" scores gained high serum IgG levels and avidity levels against these antigens after treatment (Figure 3(b)). Meanwhile, neither the IgA levels nor IgA avidity levels were found to be associated with "smear at entry" values (data not shown).

3.4. Analysis of the Relationship between Serum CRP Level and Serum Mtb Antigen-Specific Antibody. To observe the association between antibody responses and the progression of disease, the antibody levels as well as avidity levels were compared to the serum CRP levels. We found that CFP-10 and MDP1 IgG avidity before treatment showed a reverse association with serum CRP levels at entry (Figure 4(a)). It was also found that HrpA-IgA level before treatment showed a negative correlation with CRP before treatment, and MDP1- and Acr-IgA avidity after treatment also showed a negative correlation with CRP after 60 days of treatment (Figure 4(b)).

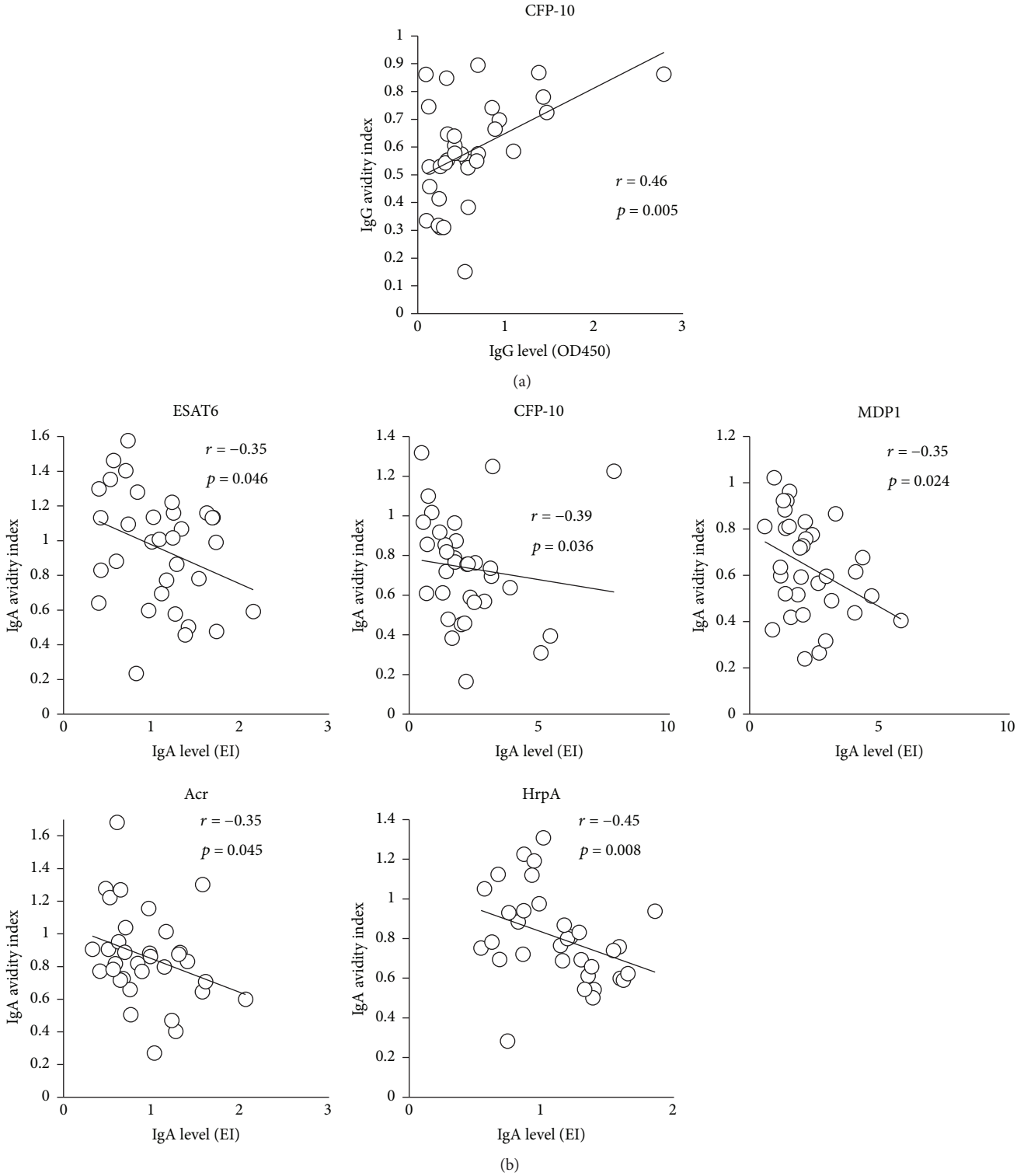


Figure 2: (a) Correlation between IgG levels and IgG avidity index against CFP-10 after treatment. (b) Correlation between IgA levels and IgA avidity index against 5 antigens after treatment.

4. Discussion

In this study, we evaluated the relationship between clinical parameters and antibody avidity to mycobacterial antigens. The aims of this study were (1) identification of the antigens that induce antibody production and maturation in TB patients and (2) evaluation of the correlation between serum antibody levels and antibody avidity to mycobacterial antigens and the host inflammatory response in TB progression. We tested 7 antigens that are highly immunogenic and are

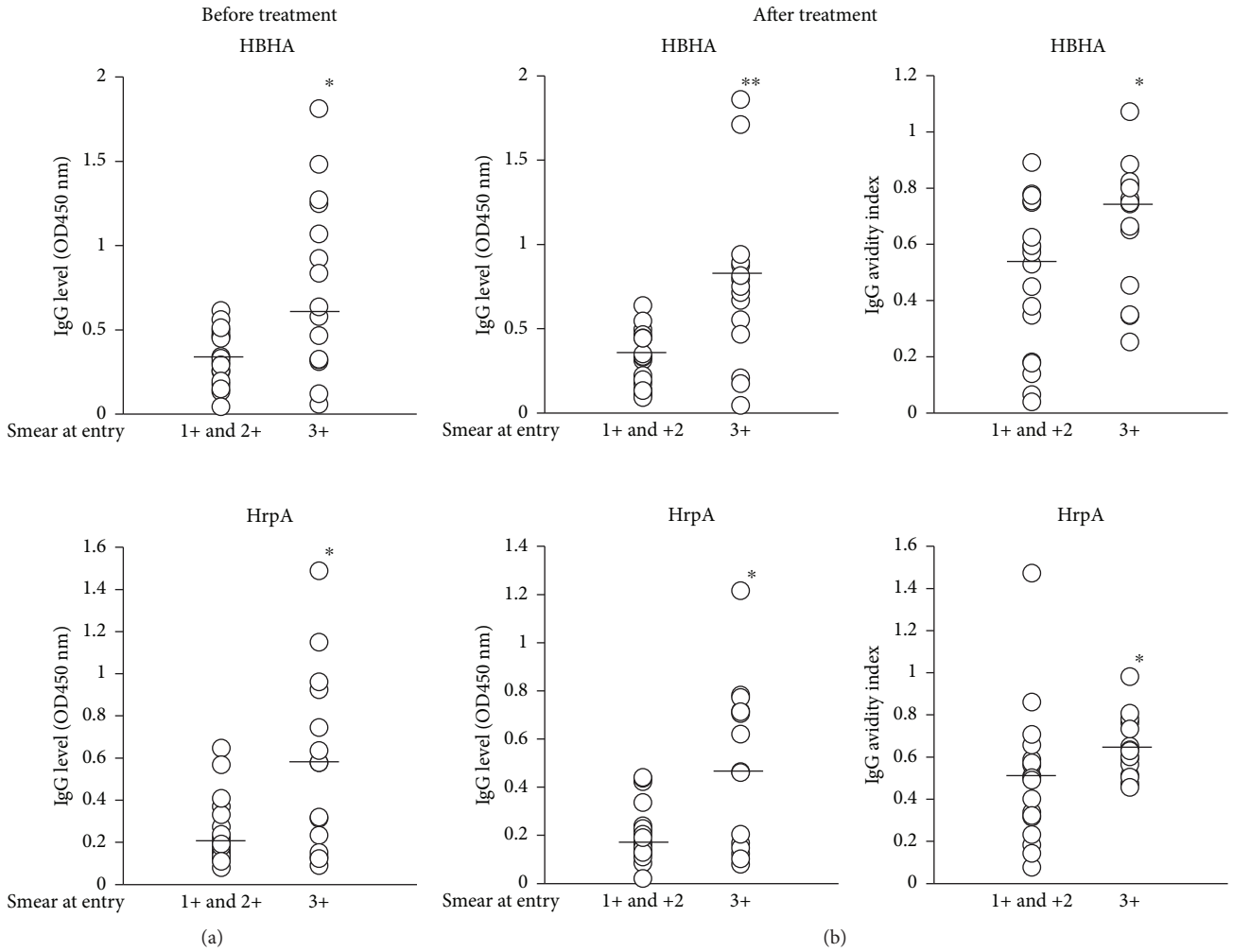


Figure 3: (a) Levels of serum IgG against HBHA and HrpA before treatment in smear at entry subgrouped between 1+ and 2+ and 3+. Vertical lines: mean values and $^{*}p < 0.05$. (b) Serum IgG levels and avidity indices against HBHA and HrpA after treatment in smear at entry subgrouped between 1+ and 2+ and 3+. Vertical lines: mean values and $^{**}p < 0.01$, $^{*}p < 0.05$.

reported to be vaccine-candidate antigens. Acr is known to be a member of the dormancy regulon-encoded antigens. HrpA, also known as Acr2, is a member of the α-crystallin family in mycobacteria, which shares 30% homology with Acr, and is strongly upregulated following infection of macrophages [41]. Both antigens have been found to contribute to prolonged Mtb infection and disease progression. Ag85A is involved in mycobacterial cell wall assembly [42] and is primarily expressed during the early stages of infection, during which Mtb cells replicate rapidly and require synthesis of cell wall components. MDP1 is a mycobacterial histone-binding protein [43], which is reported to be strongly highly expressed during latent Mtb infection [44]. HBHA is a heparin-binding hemagglutinin protein and a mycobacterial surface-expressed adhesin, which is reported to be involved in extrapulmonary dissemination of Mtb [45] by enhancing adherence and phagocytosis of mononuclear phagocytes [46].

In this study, we found a significant decrease in IgG levels against Acr and HrpA and IgG avidity to MDP1 and Ag85A during treatment. In our previous study, we showed that MDP1 and Ag85A are expressed in Mtb cells inside tuberculous granuloma lesions in an asymptomatic subject and induced production of specific antibodies in latent TB patients [44]. In contrast to IgG responses, we did not find any significant differences in IgA antibody levels and IgA avidity to Mtb antigens when compared before and after treatment.

We found that IgG avidity to CFP-10 and MDP1 before treatment exhibited a negative association with serum CRP levels at entry. As IgG avidity is generally higher in chronically infected patients than patients with acute infection, IgG avidity to these antigens may reflect the duration of Mtb infection. Additionally, we observed that levels of IgG antibodies against HBHA and HrpA exhibited a positive relationship with "smear at entry," and IgG avidity elevation during treatment was observed in patients with high IgG levels against these antigens before treatment. These findings indicate that patients with high bacterial burden produced higher levels of IgG against HBHA and HrpA. In agreement with our findings, several studies have indicated that serum IgG antibody levels against Mtb antigens correlate with the bacterial burden in active TB patients [10, 47, 48]. On the other

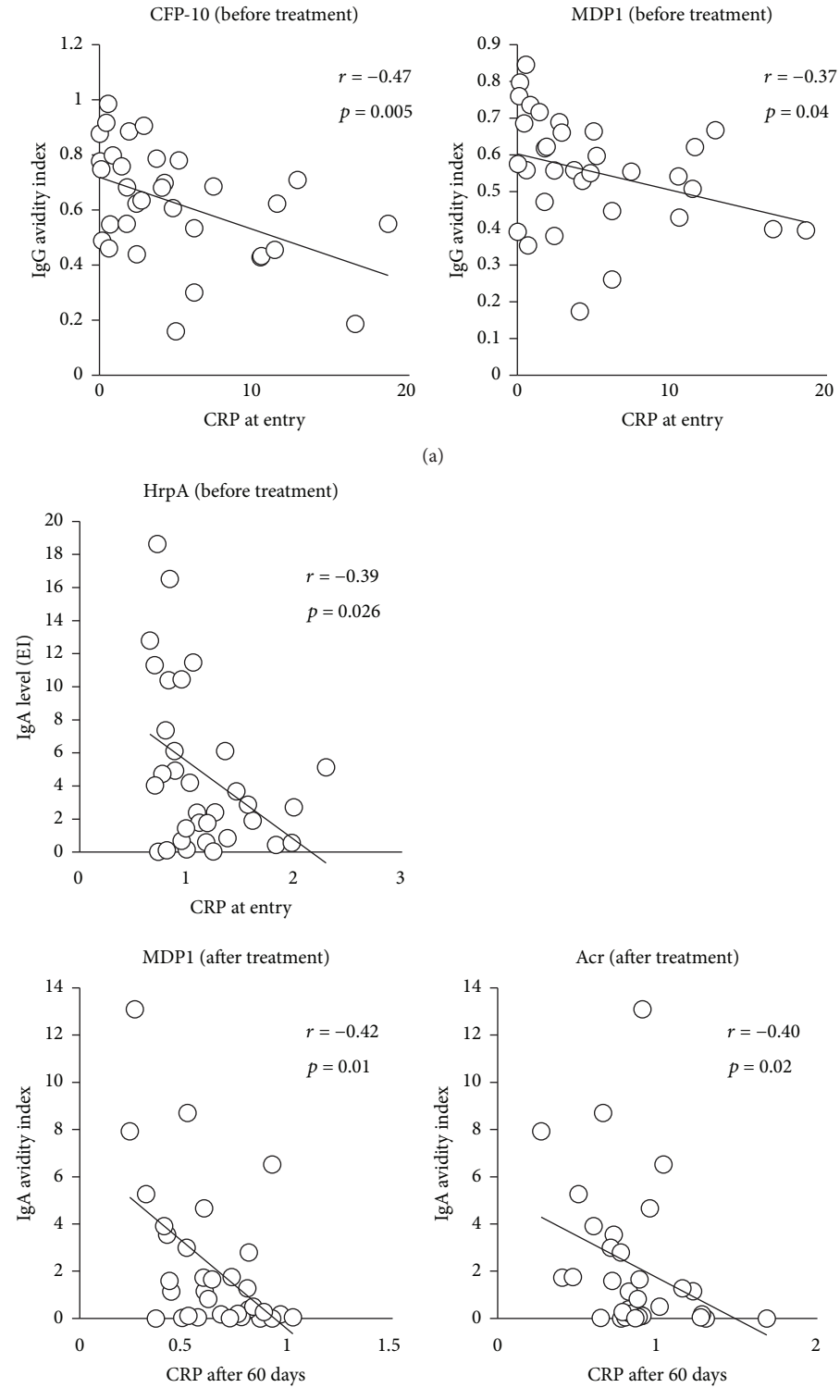


Figure 4: (a) Correlation between CRP at entry and IgG avidity index against CFP-10 and MDP1 before treatment. (b) Correlation between CRP at entry and IgA level against HrpA before treatment and correlation between CRP after 60 days and IgA avidity index against MDP1 and Acr after treatment.

hand, there were other studies showing that antibody levels do not always correspond to the bacterial load in TB patients [49, 50]. The inconsistency of these results may be due to the heterogeneity of antibody responses to Mtb antigens between individuals and among patients in different clinical stages [39, 40]. In experimental models, intravenous administration of IgG against mycobacterial lipoarabinomannan reduced bacterial load in the lung in mice intravenously infected with

Mtb [25]. Another study using a mouse TB model revealed that intranasal administration of antigen 85B-HBHA fusion protein as a booster following BCG vaccination showed a significant reduction in bacterial load [51]. On the other hand, Mtb-reactive monoclonal antibodies increased uptake of Mtb cells by human lung epithelial cells [52]. These data suggest that the protective function of IgG may depend on the ability of the target antigen to induce sufficient IgG production. In addition, a recent study demonstrated that most humans with active TB exhibited lower serum IgG avidity to the Mtb cell surface [53]. Although these published data suggest a protective role for IgG in TB, further studies are needed for the identification of protective Mtb antigen epitopes and the induction of mature B cells that produce high-avidity antibodies crucial for the development of an effective vaccine.

As IgA in its secretory form is the main effector molecule of the mucosal immune system and serves as the first line of defense against pathogen invasion initiated at mucosal surfaces, we also investigated the role of IgA in TB progression [54]. We found that bacterial load did not affect serum IgA levels or IgA avidity to Mtb antigens. On the other hand, patients with elevated IgA levels against HrpA at the initial visit were found to exhibit lower CRP levels at entry. These results suggest that early induction of high levels of specific antibodies and avidity may suppress the inflammatory response. We have now confirmed that high levels of HrpA were significantly associated with lower serum CRP levels at entry in two different settings, namely, the current longitudinal study, as well as a previous cross-sectional study [36]. Moreover, we demonstrated that induction of antibody avidity to mycobacterial antigens may be associated with lower serum CRP levels, a marker of lower inflammatory status. Consistent with our results, several reports in both animal and human models showed that IgA provides early protection against Mtb infection. In mice, IgA-deficiency leads to increased susceptibility to intranasal BCG infection [55]. Another study revealed that mouse IgA monoclonal antibody against the Mtb antigen Acr reduced early pulmonary Mtb infection in mice [56]. It was also reported that passive administration of purified secretory IgA from human colostrum reduced the pneumonic area in a murine infection model [24]. These findings indicate that IgA may play a pivotal role in the host's early defense against Mtb invasion in the respiratory tract.

Despite the great efforts that have been made to develop a novel vaccine that can effectively induce CMI to eliminate intracellular Mtb bacilli, these vaccine candidates have failed to induce better protection than BCG. As recent studies extensively demonstrated that mucosal immune responses are important in protecting the host from Mtb infection [57, 58], vaccine strategies that attempt to enhance mucosal immunity should be included in future TB vaccine development efforts. While the results of the present study suggest a possible role of specific antibodies in TB protection, and thus the benefit of potential inclusion of some of the investigated antigens in the development of a future vaccine candidate, the results are still not fully conclusive at this point, highlighting the need for further research in this area.

5. Conclusion

To our knowledge, this is the first human study to investigate the relationship between the kinetics of humoral antibody to various Mtb antigens and the clinical disease status during the treatment. The correlation between humoral immunity and bacterial and clinical parameters was analyzed for the first time. Antigen-specific IgA suppresses inflammatory responses in active tuberculosis patients. The data in this study support the inclusion of strategies that elicit humoral immunity when developing vaccines against tuberculosis.

Conflicts of Interest

The authors declare that they have no conflict of interests.

Authors' Contributions

Mamiko Niki, Yuji Miyamoto, and Yoshihiko Hoshino performed the laboratory experiments. Takashi Yoshiyama, Masao Okumura, Makoto Niki, Ken-ichi Oinuma, and Sohkichi Matsumoto contributed to the experimental system and statistical analysis. Mamiko Niki, Takashi Yoshiyama, and Yoshihiko Hoshino contributed to the planning and proposal of the work. Takashi Yoshiyama, Yukihiro Kaneko, Hideo Ogata, Hajime Goto, Shoji Kudoh, and Yoshihiko Hoshino coordinated the work.

Acknowledgments

This work was supported in part by a Grant-in-Aid for Research on Emerging and Reemerging Infectious Diseases from the Ministry of Health, Labour, and Welfare of Japan for Yoshihiko Hoshino (JP18fk0108075) and by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science for Mamiko Niki and Yoshihiko Hoshino.

Supplementary Materials

Supplementary 1. Supplemental Figure 1: (A) IgG responses to Mtb antigens. Levels of serum IgG against 7 antigens (ESAT-6, CFP-10, MDP1, Ag85A, Acr, HBHA, and HrpA) before treatment (labeled as "before") and after treatment (labeled as "after") were analyzed by ELISA. Data shown are the average of triplicate experiments. Vertical lines: mean values. (B) IgG avidity index to Mtb antigens. Levels of serum IgG avidity index against 7 antigens (ESAT-6, CFP-10, MDP1, Ag85A, Acr, HBHA, and HrpA) before treatment (labeled as "before") and after treatment (labeled as "after") were analyzed by Urea ELISA. Data shown are the average of triplicate experiments. Vertical lines: mean values, $^{**}p < 0.01$ and $^{*}p < 0.05$. Supplemental Figure 2: (A) IgA responses to Mtb antigens. Levels of serum IgA against 7 antigens (ESAT-6, CFP-10, MDP1, Ag85A, Acr, HBHA, and HrpA) before treatment (labeled as "before") and after treatment (labeled as "after") were analyzed by ELISA. Data shown are the average of triplicate experiments. Vertical lines: mean values. (B) IgA avidity index to Mtb antigens. Levels of serum IgA avidity index against 7 antigens (ESAT-6,

CFP-10, MDP1, Ag85A, Acr, HBHA, and HrpA) before treatment (labeled as "before") and after treatment (labeled as "after") were analyzed by Urea ELISA. Data shown are the average of triplicate experiments. Vertical lines: mean values.

Supplementary 2. Supplemental Table 1: antibody responses in individuals.

References

[1] A.-L. Bañuls, A. Sanou, N. T. Van Anh, and S. Godreuil, "*Mycobacterium tuberculosis*: ecology and evolution of a human bacterium," *Journal of Medical Microbiology*, vol. 64, no. 11, pp. 1261–1269, 2015.

[2] T. H. M. Ottenhoff and S. H. E. Kaufmann, "Vaccines against tuberculosis: where are we and where do we need to go?," *PLoS Pathogens*, vol. 8, no. 5, article e1002607, 2012.

[3] A. M. Cooper, "Cell-mediated immune responses in tuberculosis," *Annual Review of Immunology*, vol. 27, no. 1, pp. 393–422, 2009.

[4] R. A. Seder and A. V. S. Hill, "Vaccines against intracellular infections requiring cellular immunity," *Nature*, vol. 406, no. 6797, pp. 793–798, 2000.

[5] N. P. Goonetilleke, H. McShane, C. M. Hannan, R. J. Anderson, R. H. Brookes, and A. V. S. Hill, "Enhanced immunogenicity and protective efficacy against *Mycobacterium tuberculosis* of bacille Calmette-Guérin vaccine using mucosal administration and boosting with a recombinant modified vaccinia virus Ankara," *Journal of Immunology*, vol. 171, no. 3, pp. 1602–1609, 2003.

[6] H. McShane, A. A. Pathan, C. R. Sander et al., "Recombinant modified vaccinia virus Ankara expressing antigen 85A boosts BCG-primed and naturally acquired antimycobacterial immunity in humans," *Nature Medicine*, vol. 10, no. 11, pp. 1240–1244, 2004.

[7] M. D. Tameris, M. Hatherill, B. S. Landry et al., "Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial," *The Lancet*, vol. 381, no. 9871, pp. 1021–1028, 2013.

[8] R. Sun, Y. A. W. Skeiky, A. Izzo et al., "Novel recombinant BCG expressing perfringolysin O and the over-expression of key immunodominant antigens; pre-clinical characterization, safety and protection against challenge with *Mycobacterium tuberculosis*," *Vaccine*, vol. 27, no. 33, pp. 4412–4423, 2009.

[9] D. F. Hoft, A. Blazevic, A. Selimovic et al., "Safety and immunogenicity of the recombinant BCG vaccine AERAS-422 in healthy BCG-naïve adults: a randomized, active-controlled, first-in-human phase 1 trial," *eBioMedicine*, vol. 7, pp. 278–286, 2016.

[10] J. M. Achkar, E. Jenny-Avital, X. Yu et al., "Antibodies against immunodominant antigens of *Mycobacterium tuberculosis* in subjects with suspected tuberculosis in the United States compared by HIV status," *Clinical and Vaccine Immunology*, vol. 17, no. 3, pp. 384–392, 2010.

[11] K. R. Steingart, N. Dendukuri, M. Henry et al., "Performance of purified antigens for serodiagnosis of pulmonary tuberculosis: a meta-analysis," *Clinical and Vaccine Immunology*, vol. 16, no. 2, pp. 260–276, 2009.

[12] S. J. Culkin, T. Rhinehart-Jones, and K. L. Elkins, "A novel role for B cells in early protective immunity to an intracellular pathogen, Francisella tularensis strain LVS," *Journal of Immunology*, vol. 158, pp. 3277–3284, 1997.

[13] J. Langhorne, C. Cross, E. Seixas, C. Li, and T. von der Weid, "A role for B cells in the development of T cell helper function in a malaria infection in mice," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 4, pp. 1730–1734, 1998.

[14] P. Mastroeni, C. Simmons, R. Fowler, C. E. Hormaeche, and G. Dougan, "*Igh-6*$^{-/-}$ (B-cell-deficient) mice fail to mount solid acquired resistance to oral challenge with virulent *Salmonella enterica* serovar typhimurium and show impaired Th1 T-cell responses to *Salmonella* antigens," *Infection and Immunity*, vol. 68, no. 1, pp. 46–53, 2000.

[15] A. Casadevall, "Antibody-mediated immunity against intracellular pathogens: two-dimensional thinking comes full circle," *Infection and Immunity*, vol. 71, no. 8, pp. 4225–4228, 2003.

[16] L. Kozakiewicz, J. Phuah, J. Flynn, and J. Chan, "The role of B cells and humoral immunity in *Mycobacterium tuberculosis* infection," *Advances in Experimental Medicine and Biology*, vol. 783, pp. 225–250, 2013.

[17] R. Wahid, S. J. Zafar, M. A. McArthur, M. F. Pasetti, M. M. Levine, and M. B. Sztein, "Live oral *Salmonella enterica* serovar Typhi vaccines Ty21a and CVD 909 induce opsonophagocytic functional antibodies in humans that cross-react with *S.* Paratyphi A and *S.* Paratyphi B," *Clinical and Vaccine Immunology*, vol. 21, no. 3, pp. 427–434, 2014.

[18] F. Bowe, E. C. Lavelle, E. A. McNeela et al., "Mucosal vaccination against serogroup B meningococci: induction of bactericidal antibodies and cellular immunity following intranasal immunization with NadA of *Neisseria meningitidis* and mutants of *Escherichia coli* heat-labile enterotoxin," *Infection and Immunity*, vol. 72, no. 7, pp. 4052–4060, 2004.

[19] S. A. Khan, C. Waugh, G. Rawlinson et al., "Vaccination of koalas (*Phascolarctos cinereus*) with a recombinant chlamydial major outer membrane protein adjuvanted with poly I:C, a host defense peptide and polyphosphazine, elicits strong and long lasting cellular and humoral immune responses," *Vaccine*, vol. 32, no. 44, pp. 5781–5786, 2014.

[20] M. Mizutani, M. Iyori, A. M. Blagborough et al., "Baculovirus-vectored multistage *Plasmodium vivax* vaccine induces both protective and transmission-blocking immunities against transgenic rodent malaria parasites," *Infection and Immunity*, vol. 82, no. 10, pp. 4348–4357, 2014.

[21] D. B. Rawool, C. Bitsaktsis, Y. Li et al., "Utilization of Fc receptors as a mucosal vaccine strategy against an intracellular bacterium, *Francisella tularensis*," *Journal of Immunology*, vol. 180, no. 8, pp. 5548–5557, 2008.

[22] J. M. Achkar and A. Casadevall, "Antibody-mediated immunity against tuberculosis: implications for vaccine development," *Cell Host & Microbe*, vol. 13, no. 3, pp. 250–262, 2013.

[23] N. Olivares, B. Marquina, D. Mata-Espinoza et al., "The protective effect of immunoglobulin in murine tuberculosis is dependent on IgG glycosylation," *Pathogens and Disease*, vol. 69, no. 3, pp. 176–183, 2013.

[24] N. Alvarez, O. Otero, F. Camacho et al., "Passive administration of purified secretory IgA from human colostrum induces protection against *Mycobacterium tuberculosis* in a murine model of progressive pulmonary infection," *BMC Immunology*, vol. 14, article S3, Supplement 1, 2013.

[25] B. Hamasur, M. Haile, A. Pawlowski, U. Schroder, G. Kallenius, and S. B. Svenson, "A mycobacterial lipoarabinomannan specific monoclonal antibody and its $F(ab')_2$ fragment prolong

survival of mice infected with *Mycobacterium tuberculosis*," *Clinical & Experimental Immunology*, vol. 138, no. 1, pp. 30–38, 2004.

[26] P. B. Kang, A. K. Azad, J. B. Torrelles et al., "The human macrophage mannose receptor directs *Mycobacterium tuberculosis* lipoarabinomannan-mediated phagosome biogenesis," *The Journal of Experimental Medicine*, vol. 202, no. 7, pp. 987–999, 2005.

[27] M. Baccard-Longere, F. Freymuth, D. Cointe, J. M. Seigneurin, and L. Grangeot-Keros, "Multicenter evaluation of a rapid and convenient method for determination of cytomegalovirus immunoglobulin G avidity," *Clinical and Diagnostic Laboratory Immunology*, vol. 8, no. 2, pp. 429–431, 2001.

[28] K. N. Ward, D. J. Turner, X. C. Parada, and A. D. Thiruchelvam, "Use of immunoglobulin G antibody avidity for differentiation of primary human herpesvirus 6 and 7 infections," *Journal of Clinical Microbiology*, vol. 39, no. 3, pp. 959–963, 2001.

[29] W. R. Usinger and A. H. Lucas, "Avidity as a determinant of the protective efficacy of human antibodies to pneumococcal capsular polysaccharides," *Infection and Immunity*, vol. 67, no. 5, pp. 2366–2370, 1999.

[30] M. M. Alam, M. Arifuzzaman, S. M. Ahmad et al., "Study of avidity of antigen-specific antibody as a means of understanding development of long-term immunological memory after *Vibrio cholerae* O1 infection," *Clinical and Vaccine Immunology*, vol. 20, no. 1, pp. 17–23, 2013.

[31] L. M. Arias-Bouda, S. Kuijper, A. Van der Werf, L. N. Nguyen, H. M. Jansen, and A. H. Kolk, "Changes in avidity and level of immunoglobulin G antibodies to *Mycobacterium tuberculosis* in sera of patients undergoing treatment for pulmonary tuberculosis," *Clinical and Diagnostic Laboratory Immunology*, vol. 10, no. 4, pp. 702–709, 2003.

[32] M. Boxus, L. Lockman, M. Fochesato, C. Lorin, F. Thomas, and S. L. Giannini, "Antibody avidity measurements in recipients of *Cervarix*® vaccine following a two-dose schedule or a three-dose schedule," *Vaccine*, vol. 32, no. 26, pp. 3232–3236, 2014.

[33] C. A. Siegrist, M. Pihlgren, C. Tougne et al., "Co-administration of CpG oligonucleotides enhances the late affinity maturation process of human anti-hepatitis B vaccine response," *Vaccine*, vol. 23, no. 5, pp. 615–622, 2004.

[34] C. L. Vermont, H. H. van Dijken, C. J. van Limpt, R. de Groot, L. van Alphen, and G. P. van Den Dobbelsteen, "Antibody avidity and immunoglobulin G isotype distribution following immunization with a monovalent meningococcal B outer membrane vesicle vaccine," *Infection and Immunity*, vol. 70, no. 2, pp. 584–590, 2002.

[35] Y. Schlesinger, D. M. Granoff, T. V. Murphy et al., "Avidity and bactericidal activity of antibody elicited by different Haemophilus influenzae type b conjugate vaccines," *JAMA*, vol. 267, no. 11, pp. 1489–1494, 1992.

[36] M. Niki, M. Suzukawa, S. Akashi et al., "Evaluation of humoral immunity to *Mycobacterium tuberculosis*-specific antigens for correlation with clinical status and effective vaccine development," *Journal of Immunology Research*, vol. 2015, Article ID 527395, 13 pages, 2015.

[37] K. Sasaki, "Molecular cytogenetic analysis of cancer of the digestive system and its clinical application," *Nihon Naika Gakkai Zasshi*, vol. 91, pp. 94–99, 2002.

[38] M. Legesse, G. Ameni, G. Medhin et al., "IgA response to ESAT-6/CFP-10 and Rv2031 antigens varies in patients with culture-confirmed pulmonary tuberculosis, healthy *Mycobacterium tuberculosis*–infected and non-infected individuals in a tuberculosis endemic setting, Ethiopia," *Scandinavian Journal of Immunology*, vol. 78, no. 3, pp. 266–274, 2013.

[39] K. Lyashchenko, R. Colangeli, M. Houde, H. Al Jahdali, D. Menzies, and M. L. Gennaro, "Heterogeneous antibody responses in tuberculosis," *Infection and Immunity*, vol. 66, no. 8, pp. 3936–3940, 1998.

[40] U. Demkow, M. Filewska, D. Michalowska-Mitczuk et al., "Heterogeneity of antibody response to myobacterial antigens in different clinical manifestations of pulmonary tuberculosis," *Journal of Physiology and Pharmacology*, vol. 58, Supplement 5, pp. 117–127, 2007.

[41] G. R. Stewart, S. M. Newton, K. A. Wilkinson et al., "The stress-responsive chaperone α-crystallin 2 is required for pathogenesis of *Mycobacterium tuberculosis*," *Molecular Microbiology*, vol. 55, no. 4, pp. 1127–1137, 2005.

[42] J. T. Belisle, V. D. Vissa, T. Sievert, K. Takayama, P. J. Brennan, and G. S. Besra, "Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis," *Science*, vol. 276, no. 5317, pp. 1420–1422, 1997.

[43] M. Furugen, S. Matsumoto, T. Matsuo, M. Matsumoto, and T. Yamada, "Identification of the mycobacterial DNA-binding protein 1 region which suppresses transcription in vitro," *Microbial Pathogenesis*, vol. 30, no. 3, pp. 129–138, 2001.

[44] M. Osada-Oka, Y. Tateishi, Y. Hirayama et al., "Antigen 85A and mycobacterial DNA-binding protein 1 are targets of immunoglobulin G in individuals with past tuberculosis," *Microbiology and Immunology*, vol. 57, no. 1, pp. 30–37, 2013.

[45] K. Pethe, S. Alonso, F. Biet et al., "The heparin-binding haemagglutinin of *M. tuberculosis* is required for extrapulmonary dissemination," *Nature*, vol. 412, no. 6843, pp. 190–194, 2001.

[46] S. L. Mueller-Ortiz, A. R. Wanger, and S. J. Norris, "Mycobacterial protein HbhA binds human complement component C3," *Infection and Immunity*, vol. 69, no. 12, pp. 7501–7511, 2001.

[47] S. Kunnath-Velayudhan, A. L. Davidow, H. Y. Wang et al., "Proteome-scale antibody responses and outcome of *Mycobacterium tuberculosis* infection in nonhuman primates and in tuberculosis patients," *The Journal of Infectious Diseases*, vol. 206, no. 5, pp. 697–705, 2012.

[48] Y. G. Hur, A. Kim, Y. A. Kang et al., "Evaluation of antigen-specific immunoglobulin g responses in pulmonary tuberculosis patients and contacts," *Journal of Clinical Microbiology*, vol. 53, no. 3, pp. 904–909, 2015.

[49] Y. Fujita, T. Doi, K. Sato, and I. Yano, "Diverse humoral immune responses and changes in IgG antibody levels against mycobacterial lipid antigens in active tuberculosis," *Microbiology*, vol. 151, no. 6, pp. 2065–2074, 2005.

[50] X. Li, H. Xu, S. Jiang et al., "TB-SA antibody test for diagnosis and monitoring treatment outcome of sputum smear negative pulmonary tuberculosis patients," *The Southeast Asian Journal of Tropical Medicine and Public Health*, vol. 42, no. 5, pp. 1147–1153, 2011.

[51] E. Stylianou, G. R. Diogo, I. Pepponi et al., "Mucosal delivery of antigen-coated nanoparticles to lungs confers protective immunity against tuberculosis infection in mice," *European Journal of Immunology*, vol. 44, no. 2, pp. 440–449, 2014.

[52] N. Zimmermann, V. Thormann, B. Hu et al., "Human isotype-dependent inhibitory antibody responses against *Mycobacterium tuberculosis*," *EMBO Molecular Medicine*, vol. 8, no. 11, pp. 1325–1339, 2016.

[53] C. C. Perley, M. Frahm, E. M. Click et al., "The human antibody response to the surface of *Mycobacterium tuberculosis*," *PLoS One*, vol. 9, no. 6, article e98938, 2014.

[54] M. E. Lamm, "Interaction of antigens and antibodies at mucosal surfaces," *Annual Review of Microbiology*, vol. 51, no. 1, pp. 311–340, 1997.

[55] A. Rodriguez, A. Tjarnlund, J. Ivanji et al., "Role of IgA in the defense against respiratory infections: IgA deficient mice exhibited increased susceptibility to intranasal infection with *Mycobacterium bovis* BCG," *Vaccine*, vol. 23, no. 20, pp. 2565–2572, 2005.

[56] A. Williams, R. Reljic, I. Naylor et al., "Passive protection with immunoglobulin A antibodies against tuberculous early infection of the lungs," *Immunology*, vol. 111, no. 3, pp. 328–333, 2004.

[57] D. Kaushal, T. W. Foreman, U. S. Gautam et al., "Mucosal vaccination with attenuated *Mycobacterium tuberculosis* induces strong central memory responses and protects against tuberculosis," *Nature Communications*, vol. 6, no. 1, p. 8533, 2015.

[58] M. Ahmed, H. Jiao, R. Domingo-Gonzalez et al., "Rationalized design of a mucosal vaccine protects against *Mycobacterium tuberculosis* challenge in mice," *Journal of Leukocyte Biology*, vol. 101, no. 6, pp. 1373–1381, 2017.

Hindawi
Journal of Immunology Research
Volume 2018, Article ID 3497401, 12 pages
https://doi.org/10.1155/2018/3497401

Research Article

Immune Response and Partial Protection against Heterologous Foot-and-Mouth Disease Virus Induced by Dendrimer Peptides in Cattle

I. Soria,[1,2] **V. Quattrocchi**,[1] **C. Langellotti**,[1,2] **M. Pérez-Filgueira**,[1,2] **J. Pega**,[1,2] **V. Gnazzo**,[2] **S. Romera**,[1,2] **J. Schammas**,[1] **D. Bucafusco**,[1,2] **S. Di Giacomo**,[1] **Beatriz G. de la Torre**,[3] **D. Andreu**,[3] **F. Sobrino**,[4] **E. Blanco**,[5] **and P. Zamorano**[1,2,6]

[1]*Instituto de Virología, Centro de Investigaciones en Ciencias Veterinarias, Instituto Nacional de Tecnología Agropecuaria (INTA)-Castelar, Buenos Aires, Argentina*
[2]*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina*
[3]*Departament de Ciencies Experimentals i de la Salut, Universitat Pompeu Fabra, 08003 Barcelona, Spain*
[4]*Centro de Biología Molecular Severo Ochoa (CSIC-UAM), 28049 Madrid, Spain*
[5]*Centro de Investigación en Sanidad Animal (CISA-INIA), Valdeolmos, 28130 Madrid, Spain*
[6]*Universidad del Salvador, Buenos Aires, Argentina*

Correspondence should be addressed to I. Soria; soria.ivana@inta.gob.ar, F. Sobrino; fsobrino@cbm.csic.es, and P. Zamorano; zamorano.patricia@inta.gob.ar

Received 15 May 2017; Revised 8 November 2017; Accepted 29 November 2017; Published 18 April 2018

Academic Editor: Pedro A. Reche

Synthetic peptides mimicking protective B- and T-cell epitopes are good candidates for safer, more effective FMD vaccines. Nevertheless, previous studies of immunization with linear peptides showed that they failed to induce solid protection in cattle. Dendrimeric peptides displaying two or four copies of a peptide corresponding to the B-cell epitope VP1 [136–154] of type O FMDV (O/UKG/11/2001) linked through thioether bonds to a single copy of the T-cell epitope 3A [21–35] (termed B_2T and B_4T, resp.) afforded protection in vaccinated pigs. In this work, we show that dendrimeric peptides B_2T and B_4T can elicit specific humoral responses in cattle and confer partial protection against the challenge with a heterologous type O virus (O1/Campos/Bra/58). This protective response correlated with the induction of specific T-cells as well as with an anamnestic antibody response upon virus challenge, as shown by the detection of virus-specific antibody-secreting cells (ASC) in lymphoid tissues distal from the inoculation point.

1. Background

The foot-and-mouth-disease virus (FMDV) causes a highly contagious disease with high morbidity in cloven-hoofed animals, including cattle and swine. FMDV can be controlled by the use of a chemically inactivated whole-virus vaccine; however, some disadvantages are associated with the use of inactivated vaccine. For example, the vaccine provides short-term protection, resulting in the need for revaccination [1], and there is a risk of the infectious virus being released during vaccine production. Therefore, a number of countries with large livestock industries have abandoned vaccination. However, this policy leaves livestock herds prone to sudden outbreaks of FMD, with dramatic effects on livestock

economy and animal welfare, as seen in the United Kingdom in 2001 [2, 3] and in turn has led to intensive research on alternative vaccination strategies.

The FMD viral particle consists of a positive-strand RNA genome, a single open reading frame (ORF) which encodes four capsid proteins, VP1, VP2, VP3, and VP4, and eleven different mature nonstructural proteins (NSP).

The B-cell binding site located in the G-H loop (around residues 140–160) of FMDV VP1 protein has been identified as a predominant epitope that elicits neutralizing antibodies against this virus in natural hosts and animal models [4, 5]. A T-cell epitope, located at residues 21 to 35 of FMDV NSP 3A, efficiently stimulates lymphocytes from pigs infected with a type C virus [6].

The current inactivated FMD vaccines only promote serological protection against a given FMDV serotype, do not confer interserotype protection, and may not, in some cases, confer intraserotype protection given the antigenic variation existing within some serotypes [7]. Additionally, these vaccines present other shortcomings, such as possible incomplete inactivation of virus, need for biosafety level 4 (BSL-4 OIE) laboratories, and requirement for a cold chain to preserve virus stability. On the other hand, the vaccine virus must be purified enough as not to induce detectable antibodies against viral NSP to allow a distinction between vaccinated and infected animals [8].

Peptide vaccines are an attractive alternative strategy that relies on the usage of short peptide fragments to engineer the induction of highly targeted immune responses, consequently avoiding allergenic and/or reactogenic sequences [9]. Various synthetic peptide or recombinant protein vaccines based on the FMDV VP1 G-H loop have been shown effective in pigs [10–12], but they have shown limited efficacy in cattle [13–15], pointing to the limitations of these vaccines in eliciting broad protective responses in different hosts. Synthetic peptides are particularly attractive FMDV vaccine candidates as they are highly pure, defined, stable, and safe, and due to their modular approach, they can incorporate different B- and T-cell peptides [9, 16].

Multiple antigenic peptides (MAPs) are dendrimeric (branched) macromolecules built from a lysine core from which a defined number of epitopes radiate [17, 18]. An effective peptide vaccine needs a B-cell epitope to elicit a high neutralizing antibody response and a T-cell epitope to provide adequate cooperation between T-cells and B-lymphocytes.

The dendrimeric peptide design improves the effectiveness of viral antigenic site presentation to the immune system. Recent studies indicate that vaccination with dendrimeric peptides based on the amino acid sequence of 3A (T-cell epitope) and VP1 GH loop (B-cell epitope) from the type O FMDV O/UKG/11/2001, and branched by means of thioether or maleimide conjugation chemistries, elicits an immune response that achieved protection in up to 100% of the vaccinated pigs [16]. Likewise, we recently reported that similar dendrimeric peptides, based on the amino acid sequences from the type O FMDV O1/Campos/Bra/58, including a VP4 sequence as T-cell epitope, can protect cattle against homologous challenge [19].

The aim of this study was to investigate whether dendrimeric peptides elicited protection against heterologous viruses, a relevant issue for efficient vaccine design. To this end, the immune response elicited in cattle by dendrimers containing amino acid sequences of 3A and VP1 GH loop from type O FMDV O/UKG/11/2001, B_2T and B_4T, and the protection they afforded against the heterologous type O virus O1/Campos/Bra/58, was analyzed.

Our results indicate that B_2T and B_4T elicited specific humoral responses in cattle and conferred partial protection against the challenge with a heterologous virus O1/Campos/Bra/58. This protective response correlated with the induction of FMDV-specific T-cells as well as with an anamnestic antibody response upon virus challenge, as shown by the detection of virus-specific ASC in lymphoid tissues distal from the inoculation point.

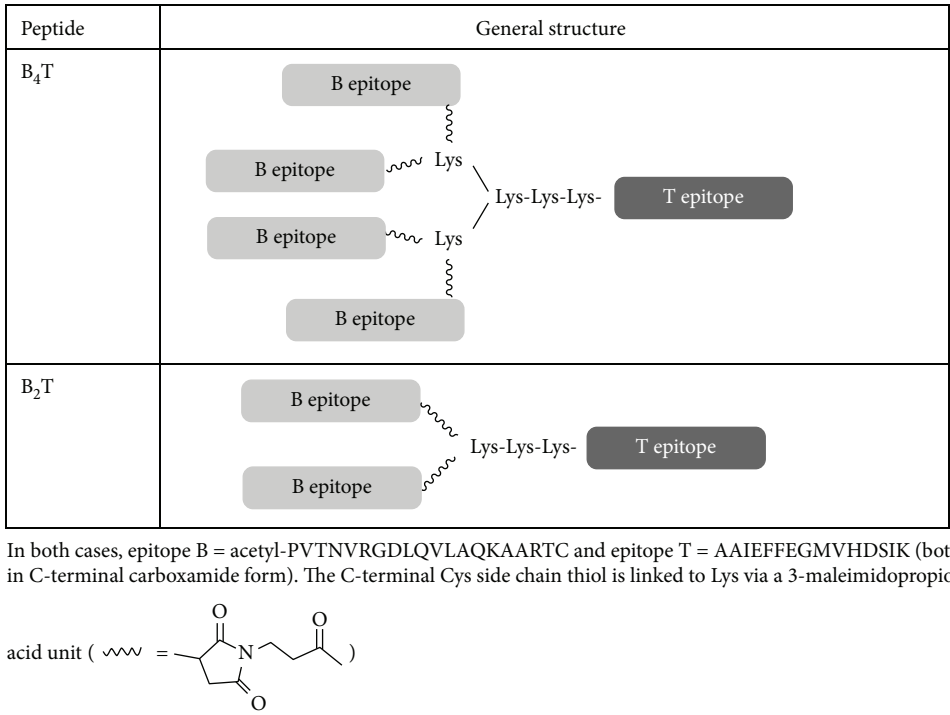
2. Material and Methods

2.1. Peptides. The dendrimeric peptides reproduced the B-cell (PVTNVRGDLQVLAQKAART, residues 136–154 of VP1) and T-cell (AAIEFFEGMVHDSIK, residues 21–35 of 3A) epitopes of FMDV O-UKG 11/01 (Figure 1). As detailed in [19], B_2T and B_4T constructions were assembled by conjugation of a T-epitope N terminally elongated with Lys residues providing 2 or 4 levels of branching and functionalized with chloroacetyl units and an N-acetylated B epitope with a C-terminal Cys whose thiol group reacts with the chloroacetyl group to give a thioether link. Additional details on the synthesis are available in previously published works [11, 20]. The final products were purified to near homogeneity by HPLC and characterized by mass spectrometry.

2.2. Virus. FMDV O1/Campos/Bra/58 was kindly provided by Biogenesis Bagó SA as binary ethylene-imine (BEI) inactivated (iFMDV). Purified virus was obtained by a sucrose density gradient centrifugation method [21] and was used for ELISA and lymphoproliferation assay. For challenging and virus neutralization assays, infective FMDV O1/Campos/Bra/58 (kindly donated by the Argentine National Service of Animal Health) was used in BSL-4 OIE laboratories and boxes at INTA. The sequence corresponding to the B-cell epitope of VP1 from FMDV O1/Campos/Bra/58 (140–158) comprises the amino acid residues ***A***V***P***NVRGDLQVLAQK***V***ART. The amino acids that differ between strains O1/Campos/Bra/58 and O/UKG/11/2001 are those corresponding to positions 140, 142, and 156 (indicated with the italic and bold formats).

A virus stock derived from FMDV isolated O/UKG/11/2001 (The Pirbright Institute, UK) by two amplifications in swine kidney cells was used in the virus neutralization assays.

2.3. Animals, Vaccines, Immunization, Infection, and Sampling of Cattle. Ten Hereford calves serologically negative for FMDV, approximately 6 months old, were used in the experiment. Groups of four animals were inoculated twice (days 0 and 18), by subcutaneous injection in the front left quarter, with 2 mg of B_2T or B_4T peptide in 2 ml of a water-in-oil single emulsion. The adjuvant included was the



In both cases, epitope B = acetyl-PVTNVRGDLQVLAQKAARTC and epitope T = AAIEFFEGMVHDSIK (both in C-terminal carboxamide form). The C-terminal Cys side chain thiol is linked to Lys via a 3-maleimidopropionic

Figure 1: Dendrimeric peptides used in this study.

same contained in commercial vaccines. At 38 days post-vaccination (dpv), the animals were challenged by nasal instillation with 1 ml (0.5 ml for each nostril) of 10000 of 50% bovine infective doses (BID50) of infective FMDV O1/Campos/Bra/58 (determined by titration on cattle tongue) [22–26]. This method is intended to mimic a natural FMDV infection [27]. Control unvaccinated bovines ($n = 2$) were challenged at the same time, and the same procedure was followed. All animals were monitored for 7 days for the emergence of FMD clinical signs and then were euthanized. The clinical score was determined by the number of feet presenting FMD lesions (with score one for each foot with lesions typical of FMDV) plus the presence of vesicles in the snout (score one) and/or mouth (score one), 6 being the maximum score.

Seven days postchallenge (dpc), all animals were checked for FMDV-induced lesions on the feet and tongue. Bovines with the absence of FMDV-induced lesions at the feet were considered as protected to podal generalization (PPG), while, animals with a delay in the onset of symptoms were considered partially protected (PP). At 7 dpc, different lymphoid organs were obtained postmortem from each animal: mandibular lymph nodes (ML), medial retropharyngeal lymph nodes (MRL), and tracheobronchial lymph nodes (TBL). All lymphoid organs were collected aseptically and placed in ice-cold wash buffer (RPMI 1640, 10 mM HEPES, 100 U/ml penicillin G sodium, 100 μg/ml streptomycin, and 20 μg/ml gentamicin) until processing.

Another five calves were immunized by subcutaneous injection with a single dose of commercial FMDV vaccine (water-in-oil single emulsion containing FMDV strains A Arg 2000, A Arg 2001, A24 Cruzeiro, and O1 Campos). This vaccine has been approved by the Argentine Animal Health Service (SENASA) with more than 80% of expected percentage of protection against all vaccine strains [28]. Experiments were performed according to the INTA ethics manual *Guide for the Use and Care of Experimental Animals*. The protocol was approved by the Institutional Animal Care and Use Committee (CICUAE INTA CICVyA) (Permit Number: 14/2011).

2.4. Measurement of Anti-Dendrimer and Anti-FMDV Antibodies. For the estimation of the immune response elicited by the dendrimers, we followed the methods of Soria et al. [19]. An indirect ELISA was used for anti-dendrimer antibody measurement. MaxiSorp 96-well plates (Nunc) were coated with B_4T peptide (30 μg/ml), the plates were washed and blocked with PBST-OVA 1%, and dilutions of serum samples were added. After incubation, the plates were washed and horseradish peroxidase- (HRP-) labeled goat anti-bovine IgG antibody (KPL, USA) was added. After washing, *ortho*-phenylenediamine- (OPD-) H_2O_2 was added as HRP substrate.

FMDV-specific antibodies were detected by means of an indirect ELISA, as described by Quattrocchi et al. [29]. Briefly, Immulon II 96-well ELISA plates were coated with 2.6 μg/ml FMDV O1/Campos/Bra/58 and processed as described above.

The antiviral ELISA detailed above was modified in order to detect FMDV-specific IgG1 and IgG2 (in sera) and IgG1 and IgA (in nasal swabs) antibodies. After incubation with samples, a mouse anti-bovine IgG1, IgG2, or IgA monoclonal antibody was added (kindly provided by Dr. S. Srikumaran, University of Nebraska, USA). Lastly, a (HRP)-labeled goat anti-mouse IgG antibody was added after wash. OPD was

used as HRP substrate. Absorbance was recorded at 492 nm (A492) in a microplate photometer (Multiskan FC, Thermo). The cut-off was established as the mean A492 of the negative sera (from all unvaccinated animals) plus two standard deviations (SD). Antibody titres were calculated for IgG1 and IgG2 as $\log_{10}$ of the last reciprocal dilution above cut-off. IgA levels were expressed as the ratio between the OD A492 of the nasal swabs from 22 dpv to 0 dpv. Positive control sera were included in every plate.

2.5. Neutralizing Index. The neutralizing index (NI) of serum (variable virus and fixed serum) from cattle immunized with B_2T, B_4T, or conventional vaccine, at 38 dpv (upon 2 doses of peptide), was measured. A 1/16 serum dilution was incubated with 10-fold dilutions of infective FMDV (1000 to 1 of 50% tissue culture infective dose—TCID50), and the infective virus recovered was determined by a TCID50 assay. The NI of a serum was calculated as the ratio between the titres of the virus in the presence of vaccinated animal serum and in the presence of a negative serum. The results were expressed as $\log_{10}$ of NI.

2.6. Neutralizing Antibody Titres. Serum samples were examined for anti-FMDV neutralizing antibodies (fixed virus and variable serum) as described before [29]. Briefly, serial dilutions (from 1/4 to 1/512) of inactivated sera were incubated for 1 h at 37°C with 100 TCID50 of infective FMDV O/UKG/11/2001 or O1/Campos/Bra/58. Then virus-serum mixtures were seed on BHK-21 monolayers. After 40 min at 37°C, fresh MEM-D/2% fetal calf serum was added to the monolayers and incubated at 37°C, under 5% CO_2. Cytopathic effects were observed after 48 h. Titres of virus neutralizing antibodies (VNT) were expressed as $\log_{10}$ of the reciprocal of the serum dilution, which neutralizes 50% of 100 TCID50 FMDV.

2.7. Lymphoproliferation Assay. Peripheral blood mononuclear cells (PBMC) were obtained from cattle as described [30]. To this end, 100 μl of 2.5×10^6 cells/ml suspension were added to 96-well plate containing (i) 5 μg/ml iFMDV; (ii) 50 μg/ml of B_2T, B_4T, or T peptides; and (iii) 5 μg/ml concanavalin A (Sigma-Aldrich, St. Louis, MO) and the cells were incubated at 37°C in 5% CO_2 atmosphere for 4 days. During the last 18 h of culture, 1 μCi [3H]-thymidine (sp. act. 20 Ci (740 Gbq)/mMol; PerkinElmer) was added to each well. Cells were collected using a semiautomatic harvester (Skatron), and the incorporation of radioactivity into the DNA was measured by liquid scintillation counting with a counter unit (Wallac 1414, PerkinElmer) that was controlled by the WinSpectral software system. Results were expressed as stimulation index (SI). The SI was calculated as the cpm of antigen-specific proliferation/cpm of cell basal proliferation (in the absence of antigen).

2.8. Interferon-Gamma Detection. PBMC were cultured with either 50 μg/ml of B_2T, B_4T, or T peptides or with 5 μg/ml iFMDV for 72 h. Supernatants were analyzed using ELISA as described previously [30]. Briefly, plates were coated with a mAb against interferon-gamma (IFN-γ) (kindly donated by Dr. L. Babiuk). Samples and recombinant IFN-γ standard (Serotec, UK) were added, and IFN was detected using rabbit polyclonal anti-IFN-γ antibodies. After incubation, biotinylated goat anti-rabbit IgG antibody was added and then HRP-conjugated streptavidin (KPL, USA) was added. The plates were washed, incubated with (OPD)-H_2O_2, and read at 492 nm. The IFN-γ concentration was calculated from interpolation of data in the standard curve.

2.9. ELISPOT Assay for FMDV-Specific ASC. Mononuclear cell (MNC) suspensions were obtained from lymphoid tissues as previously described [31]. A FMDV-ASC ELISPOT assay was developed for this study. Ninety-six-well nitrocellulose plates (Millipore, MA) were coated overnight with 2.4 μg/well inactivated purified FMDV O1/Campos/Bra/58 and blocked with 4% skim milk for 1 h at room temperature (RT). MNC were seeded in FMDV-coated plates in 2-fold dilutions (2.5×10^5 and 1.25×10^5 cells per well) in triplicate, and wells were incubated overnight at 37°C with 5% CO_2. After 5 washes with phosphate-buffered saline (PBS), mouse anti-bovine IgG1 or IgG2 monoclonal antibodies (BD-Serotec, Oxford, UK) were added (1 : 500 dilution) and incubated for 1 h at RT. Reactions were revealed with anti-mouse IgG (HRP)-labeled conjugate (KPL, UK) for 1 h at RT, followed by the addition of TrueBlue peroxidase substrate (KPL, UK). IgM and IgA ASC were detected with HRP-labeled sheep anti-bovine IgM and IgA sera (Bethyl), diluted 1 : 5.000, and revealed as described above. Spots corresponding to ASC were visualized and counted manually under a stereomicroscope. Spots from control wells were subtracted from experimental wells, and results were expressed, unless otherwise indicated, as the mean number of ASC per 1×10^6 cells for triplicate wells.

2.10. Statistical Analysis. The InfoStat program was used. One-way analysis of variance (ANOVA) and posttests were used to compare data between three or more groups.

3. Results

3.1. B_2T and B_4T Induce Anti-Peptide and FMDV-Specific Antibodies in Cattle. At 38 dpv, all animals inoculated with either B_2T or B_4T constructs developed specific and pronounced anti-peptide (Figure 2(a)) as well as anti-FMDV total IgG (Figure 2(b)) and IgG1 responses (Figure 2(c)).

At 38 dpv, high anti-FMDV IgG titres were detected in all animals with an average titre of 3.4 ± 0.4 and 3.3 ± 0.3 in B_2T and B_4T groups, respectively (Figure 2(b)). However, some animals showed a significant increase in IgG titre only after the second peptide dose (168, 164, and 166), while the others were able to achieve high IgG titres since the first immunization. Lastly, the results showed that the anti-FMDV IgG1 was the predominant isotype in all vaccinated animals given that there was a minor difference between total IgG and IgG1, and low levels of specific FMDV IgG2 were detected in B_2T and B_4T groups with average antibody titres of 1.1 ± 0.3 and 1.9 ± 0.7, respectively (data not shown).

3.2. FMDV-Specific Mucosal Immunity. Animals from the B_4T group exhibited high levels of anti-FMDV IgG1 in nasal secretions at 22 dpv, with the exception of bovine 166;

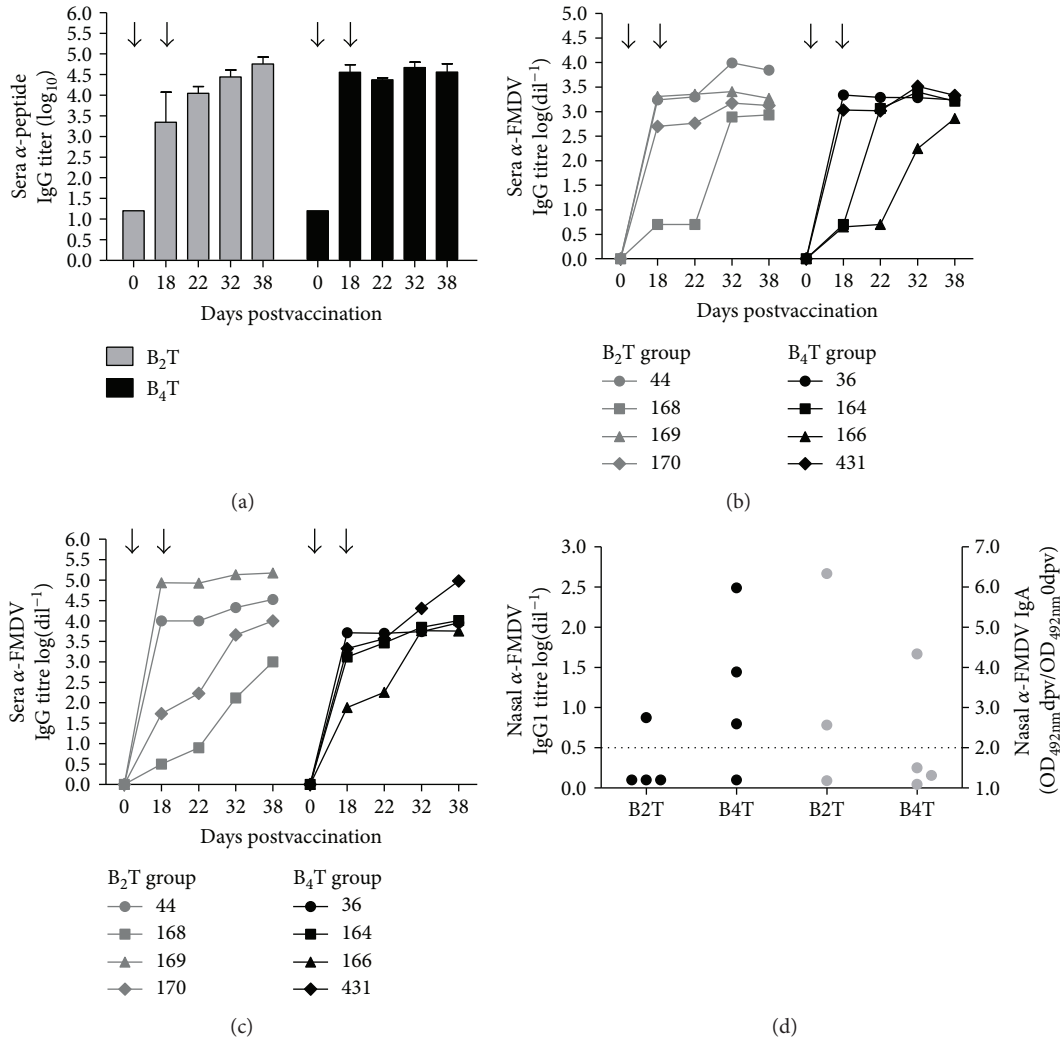


Figure 2: Antibody detection by ELISA in vaccinated cattle. Animals were immunized on days 0 and 18 (arrow) with B_2T or B_4T vaccine. (a) Kinetics of anti-peptide serum antibodies. Bars represent the mean IgG titres from bovines in each group (gray, B_2T; black, B_4T) throughout the experiment (error marks, SD). (b, c) Kinetics of total IgG and IgG1 anti-FMDV O1/Campos/Bra/58 serum antibodies. Titres were calculated as $\log_{10}$ of the last reciprocal dilution above cut-off. Data points represent the IgG titre (b) or IgG1 titre (c) from each animal represented by different shapes (right legend) throughout the experiment. (d) FMDV-specific mucosal IgG1 and IgA responses. Nasal swabs were collected at 22 dpv. Each point represents the nasal IgG1 anti-FMDV antibody titres ($\log_{10}$) (black) or IgA (gray) anti-FMDV O1/Campos/Bra/58 antibody level of each animal. The cut-off was established as the mean value of mock-vaccinated animals plus twice the SD value (dotted line).

however, at this time, animals from the B_2T group did not present high anti-FMDV IgG1 titres (Figure 2(d)). When IgA was measured in nasal secretions, animals 44 and 170 in the B_2T group and 36 in the B_4T group showed positive anti-FMDV IgA levels, indicating that these peptide constructs were able to induce not only systemic but also local mucosal immunity.

3.3. Analysis of the Neutralizing Capacity of the Sera. The VNT against the homologous virus O/UKG/11/2001 were determined at 32 dpv, and average values of 1.2 ± 0.3 and 1.3 ± 0.3 were found in the B_2T and B_4T groups, respectively (Table 1). Although the VNT against the heterologous type O virus (O1/Campos/Bra/58) was in the limit of the detection threshold, a $\log_{10}$ neutralization index with values of 1.3 ± 0.5 and 1.8 ± 0.7 could be determined for sera from animals of the B_2T and B_4T groups, respectively (Table 1). As expected, no NI values were found in the preimmune sera ($T = 0$). The $\log_{10}$ NI of sera from 4 bovines vaccinated with commercial vaccine was 2.0 ± 0.3 (data not shown).

3.4. Specific Cellular Immune Response and IFN-γ Release in Vaccinated Animals. Before challenge, at 32 dpv, specific *in vitro* lymphoproliferations were conducted using different stimuli. Significant values of proliferation ($SI \geq 2$) to the peptide used for immunization (B_2T or B_4T) were found in 2 out of 4 animals of the B_2T group and in 3 out of 4 animals of the B_4T group (Table 2(a)). Responses to dendrimers not used for immunization were similar to those achieved with the immunizing peptide while the number of animals that

Table 1: Virus neutralizing titres prechallenge.

Group	Animal no.	Neutralizing antibodies (38 dpv) VNT[a] O/UK/01	$\log_{10}$ NI O1/C
B_2T	44	1.60	2.0
	170	1.20	1.3
	168	1.10	0.8
	169	0.90	1.0
B_4T	36	1.20	1.3
	431	1.75	2.7
	164	1.10	2.0
	166	1.20	1.3
Commercial vaccine	522	—	1.8
	800	—	2.1
	809	—	1.6
	810	—	2.3
	820	—	2.8
Negative controls	167	<0.9	0.0
	997	<0.9	0.3

[a]Titre of virus-neutralizing antibody at day 38 post vaccination. O/UK/01: FMDV O/UKG/11/2001; O1/C: FMDV O1/Campos/Bra/58.

recognized the T-cell peptide alone was lower. In the B_2T group, PBMC from bovine 44 significantly proliferated in response to the T peptide, and animals 168 and 169 showed no response to any stimulus. In the B_4T group, only cells from bovine 36 proliferated when stimulated with the T peptide. PBMC from negative control animals (Table 2(a)) and from all bovines at day 0 did not respond to any peptide (data not shown). In the group immunized with the commercial vaccine, 3 animals out of 5 showed positive proliferation against iFMDV and 1 out of 5 against both dendrimers (Table 2(a)).

The levels of IFN-γ secreted in vitro by PBMC from immunized animals were also determined at 32 dpv (Table 2(b)). Positive IFN-γ responses to the immunizing peptide were found in 3 out of 4 animals of both the B_2T and B_4T groups, and the responses were similar to those induced by the dendrimers not used for immunization. In the B_2T group, only cells from bovine 44 secreted IFN-γ when they were stimulated with the T peptide, whereas 3 out of 4 animals of the B_4T group secreted IFN-γ even without stimulus (Table 2(b)).

On the other hand, bovines 169 and 166 did not secrete IFN-γ and were considered as nonresponders. PBMC from negative control animals (Table 2(b)) and from all bovines at day 0 (data not shown) did not respond to any peptide.

3.5. Different Clinical Score Protection after Challenge. Since the aim of the study was to investigate the protection afforded by the dendrimeric peptides and the infection with FMDV type O other than O1/Campos/Bra/58 was not possible at INTA, bovines were challenged with this virus, an experimental design that allows the assessment of the cross-protection conferred by the dendrimers. Thus, all animals were challenged at 44 dpv by nasal instillation with infective

Table 2: Cellular immune response of cattle 32 days postvaccination analyzed by ^{3}H-thymidine incorporation (a) and IFN-γ production ($\times 10^2$ pg/ml) (b).

(a)

Group	Animal no.	SI (cpm Ag/cpm medium) Ag-B_2T	Ag-B_4T	Ag-T	iFMDV
B_2T	44	56.0	39.6	4.0	1.1
	170	3.0	5.2	1.0	1.2
	168	1.4	1.8	1.1	1.1
	169	2.0	1.9	0.9	1.5
B_4T	36	2.5	3.6	2.3	1.4
	431	3.5	1.6	1.0	1.2
	164	4.2	7.3	1.4	1.8
	166	4.0	5.7	0.8	1.3
Commercial vaccine	522	2.9	2.5	2.0	3.2
	800	1.4	2.0	1.2	1.6
	809	1.5	1.9	1.2	4.1
	810	1.2	1.4	0.9	0.9
	820	0.7	0.7	0.7	2.8
Negative controls	167	1.0	0.7	1.0	1.0
	997	1.0	1.4	0.9	0.9

(b)

Group	Animal no.	IFN-γ ($\times 10^2$ pg/ml) Medium	Ag-B_2T	Ag-B_4T	Ag-T	iFMDV
B_2T	44	7.8	57.8	76.2	61.4	7.1
	170	14.0	32.1	30.2	12.4	11.3
	168	19.3	17.3	13.2	11.8	13.3
	169	12.3	14.0	14.6	15.0	17.9
B_4T	36	28.7	36.0	16.0	27.9	28.1
	431	32.8	15.7	54.0	34.0	35.0
	164	33.9	24.2	40.0	31.2	37.8
	166	7.9	6.0	6.5	6.8	6.8
Commercial vaccine	522	51.6	30.4	19.8	11.6	36.1
	800	7.4	7.6	7.4	7.1	7.5
	809	24.3	44.2	36.1	25.8	44.2
	810	16.9	14.3	13.7	18.0	15.9
	820	7.3	7.3	6.5	9.4	6.9
Negative controls	167	8.4	5.2	6.5	7.2	6.1
	997	9.6	12.0	14.0	13.3	14.8

(a) Lymphoproliferation of PBMC from vaccinated cattle (32 dpv) determined by ^{3}H-thymidine incorporation. Results were expressed as SI. PBMC were stimulated *in vitro* following incubation with dendrimeric peptides B_4T, B_2T or epitope T, iFMDV O1/Campos/Bra/58, or medium alone. Radioactivity was measured with b-scintillation counter. SI was calculated as cpm of each antigen specific proliferation Ag/cpm of cells basal proliferation. SI values ≥ 2.5 are considered positive. (b) IFN-γ production by PBMC after peptide stimulation as in (a). Supernatants were tested by ELISA, and the results, expressed in pg/ml, were calculated by interpolation in a cytokine standard curve. For each peptide, the cut-off was calculated as the mean IFN-γ production of PBMC from animals at day 0 plus 2 SD ($\geq 15.0 \times 10^2$ pg/ml). Positive IFN-γ productions above cut-off are underlined.

FMDV O1/Campos/Bra/58, and protection was measured by monitoring clinical signs in animals after the challenge. As shown in Table 3, the two negative control animals showed typical FMDV lesions, while, remarkably, bovines

Table 3: Clinical scores of vaccinated cattle after challenge.

Group	Animal no.	Clinical score (dpc)[a]				Protection[b]
		2 dpc	3 dpc	4 dpc	**7 dpc**	
B_2T	44	0	0	0	**2**	PPG
	170	0	0	0	**5**	PP
	168	0	0	5	**6**	NP
	169	0	0	3	**6**	NP
B_4T	36	0	0	0	2	PPG
	431	0	0	0	**5**	PP
	164	0	0	3	**5**	NP
	166	0	5	6	**6**	NP
Negative controls	167	0	0	4	6	NP
	997	0	4	4	6	NP

[a]Clinical score was established after the challenge and was determined by the number of feet presenting FMD lesions plus the presence of vesicles in the snout and/or mouth, 6 being the maximum score. [b]Animals with no lesions on the feet were PPG. Animals with a delay in the onset of symptoms of disease were PP, and animals with lesions on their feet before 7 dpc were considered NP.

44 (from the B_2T group) and 36 (from the B_4T group) did not show any lesions on their feet along the 7 days of clinical observation and were considered as PPG, albeit they showed a single vesicle in the tongue at 7 dpc. In addition, animals 170 (from the B_2T group) and 431 (from the B_4T group) showed a delay in the onset of symptoms (PP) during the normal course of disease, and the lesions in the feet appeared on 7 dpc, while those in the mock-vaccinated animals appeared on 3 or 4 dpc. Bovines 168, 169, 164, and 166 were nonprotected (NP); they presented vesicle in the tongue, mouth, and feet. At 7 dpc, all animals showed lesions in their mouth or tongue (Table 3).

3.6. Mucosal Adaptive Antibody Responses in Peptide-Vaccinated Cattle after Nasal Infection. Animals were euthanized at 7 dpc, and the FMDV-specific mucosal immune responses were studied along the respiratory tract by means of a FMDV-ASC ELISPOT assay (FMDV-ASC ELISPOT). The results showed three profiles of responses (Figure 3(a)) according to the degree of protection (PPG, PP, or NP) observed in the animals. In general, PPG and PP bovines showed a very low number of ASC in mandibular lymph nodes (ML) and medial retropharyngeal lymph nodes (MRL), with the exception of bovine 431. Tracheobronchial lymph nodes (TBL) of animals 44 and 170 from the B_2T group (PPG and PP, resp.) and animals 36 and 431 from the B_4T group (PPG and PP, resp.) did not show secretory cells producing FMDV-specific antibodies (Figures 3(a) and 3(b)). When ASC from peptide-immunized NP animals were studied, IgM and IgG1 were the dominant isotypes of antibody detected in ML; high amounts of IgA ASC (>200 ASC/10^6 cells) were detected in animal 168. However, the other NP bovines (169, 164, and 166) presented a low number of IgA ASC (<50 ASC/10^6 cells). On the other hand, IgG2 ASC were detectable in ML at this time with values 10- to 80-fold lower than the IgG1 value. Finally, in vaccinated NP animals, high amounts of IgM and IgG1 ASC were detected in MRL (excluding bovine 164). Finally, animals 168, 169, and 166 presented a high level of total ASC in ML and MRL (Figure 3(b)). NP vaccinated (168,169, 164, and 166) and mock-vaccinated animals (167 and 997) also showed responses in TBL at the lower respiratory tract. IgM and IgG1 antibodies against FMDV were the isotypes secreted.

Animal 166, which showed delayed humoral response against virus, presented the highest number of IgG1 ASC ($>2\times10^3$ ASC/10^6 cells) in ML and MRL. Concordantly, bovine 166 was the only animal that showed high numbers of IgM and IgG1 ASC in TBL.

In mock-vaccinated animals (167 and 997), ML and MRL were the most stimulated secondary lymphoid organs at 7 dpc, IgM was the dominant isotype among the FMDV-ASC developed in these organs. In animal 167, IgG1 was the next isotype with regard to the detection level, with levels 10- to 30-fold lower than those detected in the ML of NP vaccinated bovines. When the total FMDV-ASC was calculated, PPG and PP animals presented very low numbers of ASC in comparison with NP vaccinated animals (Figure 3(b)).

3.7. IFN-γ Secretion by Mononuclear Cells (3 Days Post-FMDV Challenge). In order to determine the memory immunity induced in vaccinated animals after the challenge with the live virus, the level of IFN-γ secreted in vitro by mononuclear cells of those animals was measured (Figure 4). At 3 dpc, PPG animals (44 and 36) presented IFN-γ levels between 8 and 9.5×10^3 pg/ml in the supernatant of PBMC stimulated with iFMDV, B_2T, or B_4T peptides. On the other hand, PP animals (170 and 431) showed high levels of IFN-γ even without stimuli as also observed in PPG animal 36. Significant differences were found in ASC, ML, and MRL of PPG and PP animals compared to NP-vaccinated cattles. Surprisingly, bovine 169 presented high levels of IFN-γ.

4. Discussion

Synthetic peptides corresponding to the protective B- and T-cell epitopes can be considered good candidates for FMD vaccines as, among other advantages, they are safe and support a rational design and their production and characterization are simple. The development of successful peptide vaccines has been limited for a number of reasons, including

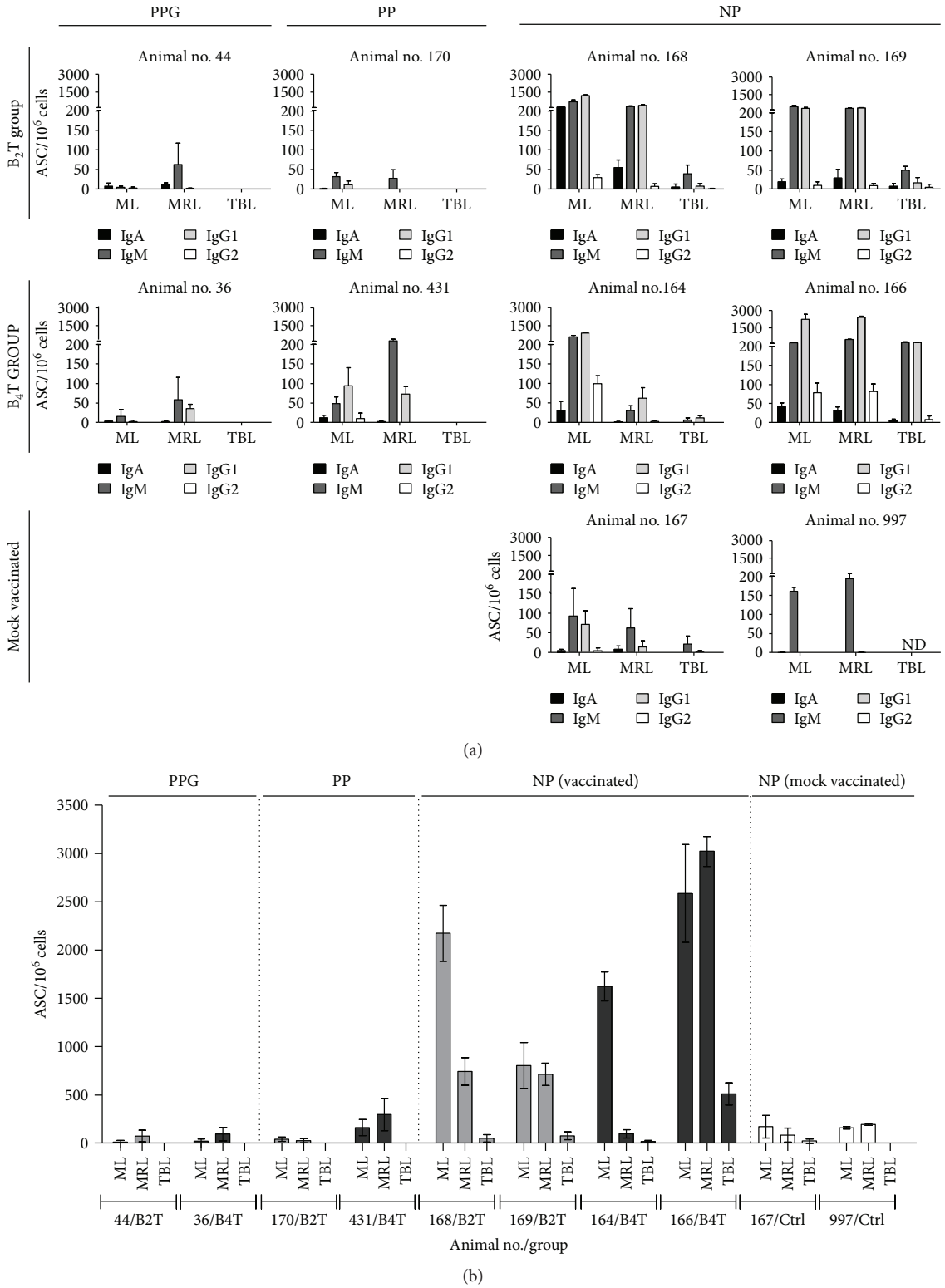


Figure 3: Profiles of the FMDV-ASC detected in B_2T- and B_4T-vaccinated cattle after FMDV challenge. (a) Mononuclear cells were purified from mandibular lymph nodes (ML), medial retropharyngeal lymph nodes (MRL), and tracheobronchial lymph nodes (TBL) and characterized by the FMDV-ASC ELISPOT assay, using monoclonal (IgG1 and IgG2) or polyclonal (IgM and IgA) antibodies against bovine immunoglobulin isotypes as probes. (b) Total FMDV-ASC in ML, MRL, or TBL. Results are expressed as the mean number of FMDV-specific ASC per 1×10^6 extracted cells. Each bar represents the mean value of 3 replicates ± SD. PPG: protected against podal generalization; PP: partial protected; NP: non-protected.

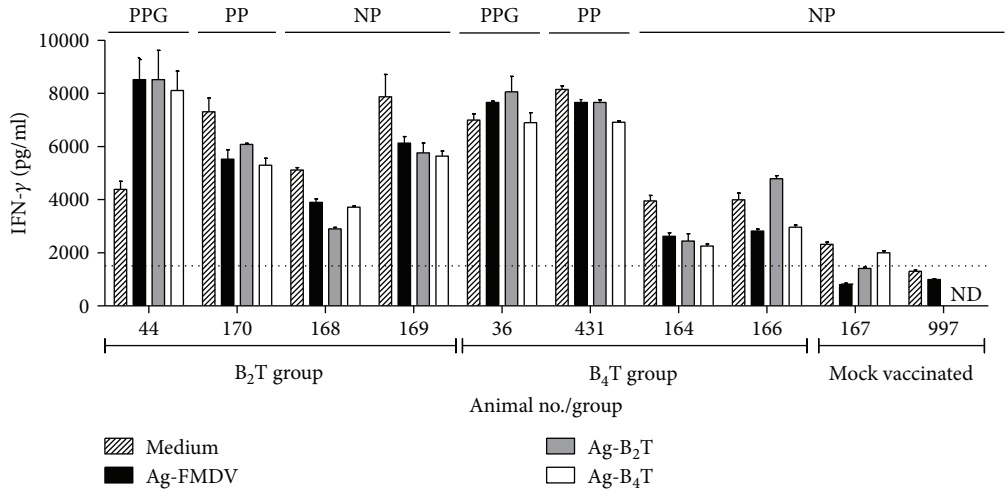


Figure 4: IFN-γ production by PBMC from peptide-vaccinated bovines after challenge. PBMC were purified at 3 dpc and cultured in the presence of peptide or inactivated virus. The supernatants were tested by a sandwich ELISA. Results are expressed in pg/ml by interpolation in a cytokine standard curve. Each bar represents the mean value of 2 replicates of supernatants ± SD. PPG: protected against podal generalization; PP: partial protected; NP: nonprotected.

those associated with "in vivo" stability, poor immunogenicity of linear peptides, and lack of adequate T-cell activation due to MHC polymorphism of the host species [32, 33]. Previous results in pigs vaccinated with B_2T or B_4T dendrimeric peptides allow concluding that multiple presentation of the B-cell epitope is advantageous over a simple juxtaposition of the epitopes for the induction of humoral and cellular immune responses [34]. Recently, Blanco et al. [16] reported that 100% of pigs vaccinated with B_2T dendrimeric peptides bearing type O FMDV O/UKG/11/2001 sequences of a B- (VP1 136–154) and a T-cell epitope (3A 21–35) were protected after the challenge with homologous FMDV. In this report, we have explored the immunogenicity of B_2T and B_4T dendrimers in cattle and showed that they can elicit cross-reactive immune responses against a heterologous type O strain, FMDV O1/Campos/Bra/58, including partial protection to challenge.

In our experiment, specific antibody responses to virus were observed in all cattle receiving peptide vaccines; however, even when neutralizing antibodies against FMDV O/UKG/11/2001 were detected, their levels were lower than those found in pigs immunized with the same peptides.

The amino acid sequence of the B-cell epitope VP1 (136–154) between FMDV O/UKG/11/2001 and O1/Campos/Bra/58, the virus used for cattle challenge, differs in 3 amino acids. Nevertheless, Wang et al. [10] reported that pigs vaccinated with a peptide containing a consensus type O VP1 sequence (residues 129–169) from 75 historic and prevalent isolates (including O1/Campos/Bra/58 and O/UKG/11/2001) and a promiscuous artificial Th site, developed humoral immunity against FMDV O1/Campos/Bra/58. Indeed, the neutralizing activity against FMDV O1/Campos/Bra/58 was detected in our peptide-immunized cattle, albeit the magnitudes of the responses were lower than those elicited against FMDV O/UKG/11/2001. B_2T and B_4T peptides induce neutralizing antibodies against FMDV O/UKG 11 at low levels, but they do not neutralize FMDV O1 Campos (100 DITC50).

Despite the presence of anti-peptide and anti-FMDV antibodies in sera, they may not have the affinity necessary to effectively neutralize the virus, and only 25% of the B_2T- or B_4T-vaccinated animals were PPG after the challenge with FMDV O1/Campos/Bra/58. It is possible that when using another adjuvant or other amounts of peptides in the vaccine, the immune response could increase in cattle, achieving the maturation of the antibodies affinity for the viral neutralization of FMDV O1/Campos/Bra/58.

The isotype of antibodies elicited in cattle by the two dendrimers differs from those induced in swine [16]. Pigs vaccinated with B_2T showed a trend towards increased levels of specific IgG1 and IgG2 relative to pigs vaccinated with B_4T. In contrast, here B_2T- or B_4T-vaccinated bovines elicit levels of IgG1 higher than those of IgG2. These changes seem to reflect marked differences in how the immune systems of swine and cattle recognize and process the dendrimeric immunogens. In any case, it is noteworthy that the same nomenclature for subclasses among different species often leads to the misleading belief that these subclasses are homologous and have the same functions.

Animal-to-animal variation is found in the protective responses evoked by peptide vaccines, including those against FMDV [13, 35], which has been associated with the MHC-restricted recognition of T-cell epitopes included in their composition. On the other hand, the different immune responses against peptide might be indicative of marked differences in the recognition of T epitopes between cattle and swine. The T-cell peptide 3A (21–35) was well studied in pigs [6] but not in cattle, and our results support that T epitope 3A (21–35) is not recognized by the majority of bovines.

Our findings suggest that in some instances, animals showing the highest immunological parameters measured against peptides and iFMDV were better protected against

viral challenge. Bovines 44 (B_2T) and 36 (B_4T) elicited high levels of antibodies against virus (although animal 36 showed levels of neutralizing antibodies of 1.2) and developed high levels of IgA specific against virus in nasal secretions as well as a positive lymphoproliferative response not only against dendrimeric peptides but also against the epitope T 3A (21–40). All these positive parameters in bovines 44 and 36 correlated with a protective immune response. Indeed, these were the only two PPG animals. On the other hand, and despite at the time of challenge the level of antibodies (measured by ELISA) being similarly high for all cattle, animals that showed modest humoral response initially (at 18 or 22 dpv) failed to be protected against viral challenge, which may be due to the lack of antibody maturation. Nonprotected animals 168 (B_2T) and 164 (B_4T) showed antibody responses against iFMDV only after receiving a second dose of vaccine, and their viral neutralization titres were lower than 1.2 (VNT positive for FMDV ≥ 1.2 [27]). These results suggest that although neutralizing antibodies are important in protecting against viral challenge, other factors could also favor protection.

Cattle are highly susceptible to FMDV, and the virus usually gains entry through the respiratory tract of these animals [36]. Moreover, FMDV replicates in tissues of the upper respiratory system [37, 38], the soft palate and pharynx being preferential sites of FMDV replication and persistence in ruminants. A feature of the mucosal system in ruminants is the prominence of IgG1 relative to IgA in nasal secretions.

The study of antibody responses in local lymphoid tissues indicates that the systemic FMD vaccination of cattle with dendrimeric peptides can effectively promote the presence of anti-FMDV ASC in lymphoid tissues associated with the respiratory tract. In addition, the detection of both FMDV O1/Campos/Bra/58-specific ASC and antibodies following vaccination shows that these peptides, encompassing FMDV O/UKG/11/2001 sequences, were able to induce a cross-reactive ASC response.

In peptide-vaccinated unprotected animals, viral challenge by nasal instillation triggered an antibody response compatible with a local anamnestic recall upon contact with replicating FMDV, suggesting that peptide vaccination might induce the circulation of virus-specific B-lymphocytes, including memory B-cells that differentiate into ASC soon after contact with the infective virus. Thus, NP animals showed a strong stimulation of FMDV-specific B-lymphocytes to locally produce antibodies all along the respiratory tract, including in the tracheobronchial lymph nodes (TBL) with frequencies of ASC much higher than those in mock-vaccinated infected animals. In the NP animals, ASC were detected in all studied organs, and the isotype of the antibodies (mainly IgM and IgG1) revealed that even when B_2T and B_4T peptides elicited specific memory B-cells, the response failed to stop the advance of the challenge virus. Conversely, in peptide-vaccinated PPG and PP animals, no FMDV-ASC were detected in TBL possibly because the virus did not reach that area. Thus, in animals 44 and 36 (PPG) and 170 and 431 (PP), cells producing antibodies against FMDV were not found in TBL, and in general the total number of ASC induced was low.

It has been proposed that structural features lend FMDV capsids towards stimulating B-cells in a T-independent manner [39, 40] and acute cytopathic viral infections can result in the accelerated induction of antibody in a T-independent manner [41, 42], providing a rapid means of stopping the systemic spread of the virus [43]. In the absence of CD4+ T-cells, cattle can produce class-switched antibody rapidly in response to the FMDV challenge [40], and a rapid induction of FMDV-specific plasma cells has been also reported in local lymphoid tissue following live-virus exposure, which, again, is consistent with a T-independent response [21]. In this report, at 7 dpc, limited amounts of IgM and IgG1 ASC were detected in ML, MRL, and TBL of one of the negative control animals (167), and only IgM was found in the other (997), a result that is consistent with a primary response against FMDV. A greater increase in the number of ASC of isotypes IgA, IgM, IgG1, and IgG2 was found by Monso et al. [20] at 6 dpc in ML, TBL, and MRL; this discrepancy may be related with the difference in the viral dose and inoculation route employed by these authors (10^7 TCID50, aerosol) compared with those of our study (10^4 BID50, nasal instillation).

Overall, our results support that immunization in cattle with dendrimeric peptides B_2T and B_4T can elicit humoral and cellular immune responses and confer partial protection against a heterologous virus challenge that is associated with the induction of solid T-cell responses as well as of an anamnestic antibody response. Experiments are in progress to address whether modifications such as the replacement of the T-cell peptide by one widely recognized by cattle can result in an improvement of the protective response elicited by these dendrimeric peptides.

Additional Points

Highlights. Dendrimeric peptides containing T and B epitopes were designed as vaccine candidates against FMDV for cattle. Dendrimeric peptides evoke in immunized cattle heterologous cellular and humoral immune responses. Tetravalent and bivalent presentation of B-cell epitopes linked to the T epitope conferred partial heterologous protection in cattle.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgments

The authors thank N. Fondevilla for his invaluable work in the BSL-4 OIE facilities at the CICVyA-INTA and J. Leiva, C. Fioroni, D. Franco, and L. Vagnoni for their help in animal handling. The authors also thank E. Rivarola for his administrative assistance and E. Smitsaard (Biogenesis Bagó) for providing the inactivated FMDV. Work at INTA was supported by the National Institute of Agricultural Technology (Grant no. PNSA 1115052) and an INTA-INIA (Spain) cooperation agreement. Work at UPF and CBMSO was supported by MINECO, Spain (Grant no. AGL2014-52395-C2).

References

[1] S. Barteling and J. Vreeswijk, "Developments in foot-and-mouth disease vaccines," *Vaccines*, vol. 9, no. 2, pp. 75–88, 1991.

[2] C. J. Laurence, "Animal welfare consequences in England and Wales of the 2001 epidemic of foot and mouth disease," *Revue Scientifique et Technique*, vol. 21, no. 3, pp. 863–868, 2002.

[3] D. K. Thompson, P. Muriel, D. Russell et al., "Economic costs of the foot and mouth disease outbreak in the United Kingdom in 2001," *Revue Scientifique et Technique*, vol. 21, no. 3, pp. 675–687, 2002.

[4] J. L. Bittle, R. A. Houghten, H. Alexander et al., "Protection against foot-and-mouth disease by immunization with a chemically synthesized peptide predicted from the viral nucleotide sequence," *Nature*, vol. 298, no. 5869, pp. 30–33, 1982.

[5] R. DiMarchi, G. Brooke, C. Gale, V. Cracknell, T. Doel, and N. Mowat, "Protection of cattle against foot-and-mouth disease by a synthetic peptide," *Science*, vol. 232, no. 4750, pp. 639–641, 1986.

[6] E. Blanco, M. Garcia-Briones, A. Sanz-Parra et al., "Identification of T-cell epitopes in nonstructural proteins of foot-and-mouth disease virus," *Journal of Virology*, vol. 75, no. 7, pp. 3164–3174, 2001.

[7] T. R. Doel, "Natural and vaccine-induced immunity to foot and mouth disease: the prospects for improved vaccines," *Revue Scientifique et Technique*, vol. 15, no. 3, pp. 883–911, 1996.

[8] Y. Cao, D. Li, Y. Fu et al., "Rational design and efficacy of a multi-epitope recombinant protein vaccine against foot-and-mouth disease virus serotype a in pigs," *Antiviral Research*, vol. 140, pp. 133–141, 2017.

[9] W. Li, M. D. Joshi, S. Singhania, K. H. Ramsey, and A. K. Murthy, "Peptide vaccine: progress and challenges," *Vaccines*, vol. 2, no. 4, pp. 515–536, 2014.

[10] C. Y. Wang, T. Y. Chang, A. M. Walfield et al., "Effective synthetic peptide vaccine for foot-and-mouth disease in swine," *Vaccines*, vol. 20, no. 19-20, pp. 2603–2610, 2002.

[11] C. Cubillos, B. G. de la Torre, A. Jakab et al., "Enhanced mucosal immunoglobulin A response and solid protection against foot-and-mouth disease virus challenge induced by a novel dendrimeric peptide," *Journal of Virology*, vol. 82, no. 14, pp. 7223–7230, 2008.

[12] J. J. Shao, C. K. Wong, T. Lin et al., "Promising multiple-epitope recombinant vaccine against foot-and-mouth disease virus type O in swine," *Clinical and Vaccine Immunology*, vol. 18, no. 1, pp. 143–149, 2011.

[13] O. Taboga, C. Tami, E. Carrillo et al., "A large-scale evaluation of peptide vaccines against foot-and-mouth disease: lack of solid protection in cattle and isolation of escape mutants," *Journal of Virology*, vol. 71, no. 4, pp. 2606–2614, 1997.

[14] L. L. Rodriguez, J. Barrera, E. Kramer, J. Lubroth, F. Brown, and W. T. Golde, "A synthetic peptide containing the consensus sequence of the G-H loop region of foot-and-mouth disease virus type-O VP1 and a promiscuous T-helper epitope induces peptide-specific antibodies but fails to protect cattle against viral challenge," *Vaccine*, vol. 21, no. 25-26, pp. 3751–3756, 2003.

[15] Z. Zhang, L. Pan, Y. Ding et al., "Efficacy of synthetic peptide candidate vaccines against serotype-A foot-and-mouth disease virus in cattle," *Applied Microbiology and Biotechnology*, vol. 99, no. 3, pp. 1389–1398, 2015.

[16] E. Blanco, B. Guerra, B. G. de la Torre et al., "Full protection of swine against foot-and-mouth disease by a bivalent B-cell epitope dendrimer peptide," *Antiviral Research*, vol. 129, pp. 74–80, 2016.

[17] J. P. Tam, Y. A. Lu, and J. L. Yang, "Antimicrobial dendrimeric peptides," *European Journal of Biochemistry*, vol. 269, no. 3, pp. 923–932, 2002, 11846794.

[18] P. M. H. Heegaard, U. Boas, and N. S. Sorensen, "Dendrimers for vaccine and immunostimulatory uses. A review," *Bioconjugate Chemistry*, vol. 21, no. 3, pp. 405–418, 2010.

[19] I. Soria, V. Quattrocchi, C. Langellotti et al., "Dendrimeric peptides can confer protection against foot-and-mouth disease virus in cattle," *PLoS One*, vol. 12, no. 9, article e0185184, 2017.

[20] M. Monso, B. G. de la Torre, E. Blanco, N. Moreno, and D. Andreu, "Influence of conjugation chemistry and B epitope orientation on the immune response of branched peptide antigens," *Bioconjugate Chemistry*, vol. 24, no. 4, pp. 578–585, 2013.

[21] J. Pega, D. Bucafusco, S. Di Giacomo et al., "Early adaptive immune responses in the respiratory tract of foot-and-mouth disease virus-infected cattle," *Journal of Virology*, vol. 87, no. 5, pp. 2489–2495, 2013.

[22] E. Maradei, J. La Torre, B. Robiolo et al., "Updating of the correlation between lpELISA titers and protection from virus challenge for the assessment of the potency of polyvalent aphtovirus vaccines in Argentina," *Vaccine*, vol. 26, no. 51, pp. 6577–6586, 2008.

[23] K. Orsel and A. Bouma, "The effect of foot-and-mouth disease (FMD) vaccination on virus transmission and the significance for the field," *The Canadian Veterinary Journal*, vol. 50, no. 10, pp. 1059–1063, 2009.

[24] K. Orsel, M. C. M. de Jong, A. Bouma, J. A. Stegeman, and A. Dekker, "Foot and mouth disease virus transmission among vaccinated pigs after exposure to virus shedding pigs," *Vaccine*, vol. 25, no. 34, pp. 6381–6391, 2007.

[25] K. Orsel, A. Dekker, A. Bouma, J. A. Stegeman, and M. C. M. de Jong, "Vaccination against foot and mouth disease reduces virus transmission in groups of calves," *Vaccine*, vol. 23, no. 41, pp. 4887–4894, 2005.

[26] N. Fondevila, D. Compaired, E. Maradei, and S. Duffy, "Validation of a real time RT-PCR assay to detect foot-and-mouth disease virus and assessment of its performance in acute infection," *Revista Argentina de Microbiología*, vol. 46, no. 3, pp. 188–195, 2014.

[27] R. F. Sellers, A. I. Donaldson, K. A. J. Herniman, and J. Paeker, "Foot-and-mouth virus," *The Lancet*, vol. 297, no. 7711, p. 1238, 1971.

[28] OIE, "Foot and Mouth Disease," *Manual of Diagnostic Tests and Vaccines for terrestrial Animals*, pp. 1–29, 2009, http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.01.08_FMD.pdf.

[29] V. Quattrocchi, J. S. Pappalardo, C. Langellotti, E. Smitsaart, N. Fondevila, and P. Zamorano, "Early protection against foot-and-mouth disease virus in cattle using an inactivated vaccine formulated with Montanide ESSAI IMS D 12802 VG PR adjuvant," *Vaccine*, vol. 32, no. 19, pp. 2167–2172, 2014.

[30] V. Quattrocchi, I. Soria, C. A. Langellotti et al., "A DNA vaccine formulated with chemical adjuvant provides partial protection against bovine herpes virus infection in cattle," *Frontiers in Immunology*, vol. 8, p. 37, 2017.

[31] J. M. Pacheco, J. Arzt, and L. L. Rodriguez, "Early events in the pathogenesis of foot-and-mouth disease in cattle after controlled aerosol exposure," *Veterinary Journal*, vol. 183, no. 1, pp. 46–53, 2010.

[32] C. Leclerc, "New approaches in vaccine development," *Comparative Immunology, Microbiology and Infectious Diseases*, vol. 26, no. 5-6, pp. 329–341, 2003.

[33] D. Rowlands, "Foot-and-mouth disease virus peptide vaccines," in *Foot-and- mouth disease: current perspectives*, pp. 335–354, Horizon Bioscience, Norfolk, UK, 2004.

[34] E. Blanco, C. Cubillos, N. Moreno et al., "B epitope multiplicity and B/T epitope orientation influence immunogenicity of foot-and-mouth disease peptide vaccines," *Clinical and Developmental Immunology*, vol. 2013, Article ID 475960, 9 pages, 2013.

[35] T. Collen, "Foot-and-mouth disease (aphthovirus): viral T cell epitopes," in *Cell Mediated Immunity in Ruminants*, B. M. Goddeeris and W. I. Morrison, Eds., CRC Press Inc., Boca Raton, FL, USA, 1994.

[36] S. Alexandersen and N. Mowat, "Foot-and-mouth disease: host range and pathogenesis," *Current Topics in Microbiology and Immunology*, vol. 288, pp. 9–42, 2005.

[37] J. W. McVicar and P. Sutmoller, "Growth of foot-and-mouth disease virus in the upper respiratory tract of non-immunized, vaccinated, and recovered cattle after intranasal inoculation," *Journal of Hygiene*, vol. 76, no. 3, pp. 467–481, 1976.

[38] Z. Zhang and S. Alexandersen, "Quantitative analysis of foot-and-mouth disease virus RNA loads in bovine tissues: implications for the site of viral persistence," *The Journal of General Virology*, vol. 85, no. 9, pp. 2567–2575, 2004.

[39] B. V. Carr, E. A. Lefevre, M. A. Windsor et al., "$CD4^+$ T-cell responses to foot-and-mouth disease virus in vaccinated cattle," *Journal of General Virology*, vol. 94, Part 1, pp. 97–107, 2013.

[40] N. Juleff, M. Windsor, E. A. Lefevre et al., "Foot-and-mouth disease virus can induce a specific and rapid $CD4^+$ T-cell-independent neutralizing and isotype class-switched antibody response in naive cattle," *Journal of Virology*, vol. 83, no. 8, pp. 3626–3636, 2009.

[41] T. Fehr, M. F. Bachmann, H. Bluethmann, H. Kikutani, H. Hengartner, and R. M. Zinkernagel, "T-independent activation of B cells by vesicular stomatitis virus: no evidence for the need of a second signal," *Cellular Immunology*, vol. 168, no. 2, pp. 184–192, 1996.

[42] B. O. Lee, J. Rangel-Moreno, J. E. Moyron-Quiroz et al., "CD4 T cell-independent antibody response promotes resolution of primary influenza infection and helps to prevent reinfection," *Journal of Immunology*, vol. 175, no. 9, pp. 5827–5838, 2005.

[43] M. F. Bachmann and R. M. Zinkernagel, "Neutralizing antiviral B cell responses," *Annual Review of Immunology*, vol. 15, no. 1, pp. 235–270, 1997.

Hindawi
Journal of Immunology Research
Volume 2018, Article ID 5463879, 13 pages
https://doi.org/10.1155/2018/5463879

Research Article

Immature Dendritic Cell Therapy Confers Durable Immune Modulation in an Antigen-Dependent and Antigen-Independent Manner in Nonobese Diabetic Mice

Jeannette Lo, Chang-Qing Xia, Ruihua Peng, and Michael J. Clare-Salzler

Department of Pathology, Immunology and Laboratory Medicine, Center for Immunology and Transplantation, University of Florida, Gainesville, FL 32610, USA

Correspondence should be addressed to Michael J. Clare-Salzler; salzler@ufl.edu

Received 17 July 2017; Revised 19 October 2017; Accepted 27 November 2017; Published 14 February 2018

Academic Editor: Yoshihiko Hoshino

Dendritic cell (DC) immunotherapy has been effective for prevention of type 1 diabetes (T1D) in NOD mice but fails to protect if initiated after active autoimmunity. As autoreactivity expands inter- and intramolecularly during disease progression, we investigated whether DCs unpulsed or pulsed with β cell antigenic dominant determinants (DD), subdominant determinants (SD), and ignored determinants (ID) could prevent T1D in mice with advanced insulitis. We found that diabetes was significantly delayed by DC therapy. Of interest, DCs pulsed with SD or ID appeared to provide better protection. T lymphocytes from DC-treated mice acquired spontaneous proliferating capability during *in vitro* culture, which could be largely eliminated by IL-2 neutralizing antibodies. This trend maintained even 29 weeks after discontinuing DC therapy and appeared antigen-independent. Furthermore, CD4+Foxp3+ T regulatory cells (Tregs) from DC-treated mice proliferated more actively *in vitro* compared to the controls, and Tregs from DC-treated mice showed significantly enhanced immunosuppressive activities in contrast to those from the controls. Our study demonstrates that DC therapy leads to long-lasting immunomodulatory effects in an antigen-dependent and antigen-independent manner and provides evidence for peptide-based intervention during a clinically relevant window to guide DC-based immunotherapy for autoimmune diabetes.

1. Introduction

Type 1 diabetes (T1D) is an autoimmune disorder resulting from the loss of self-tolerance to pancreatic islet β cell autoantigens. Efforts to redirect the immune response toward tolerance through peptide or whole autoantigen-based therapy have been shown to be effective in autoimmune mouse models, but have met with considerable setbacks in human studies [1–8]. Difficulties in translating the appropriate tolerizing antigen dose combined with the risk of activating or enhancing autoimmunity have delayed the development of antigen-specific therapy for tolerance induction into the clinical setting. Furthermore, it is uncertain whether the delivery of antigen to an already impaired immune system [9–11] is able to correct the autoimmunity.

Dendritic cell therapy provides an alternative way of delivering antigen by using ex vivo-generated cells engineered to control the direction of the immune response toward a preloaded autoantigenic peptides of interest. We and others have demonstrated that peptide-pulsed immature dendritic cell (DC) therapy prevents T1D in NOD mice, the autoimmune diabetes mouse model, when applied during the early stages of autoimmunity [12, 13]. Interestingly, protection from unpulsed DC therapy has also been reported [14–18], challenging the need for antigen. Whether these protective DCs pick up autoantigen *in vivo* or exert antigen-independent influences to the immune repertoire is unknown as most studies using DC therapy have only assessed antigen-specific changes. The global effect that DC therapy may have on nontarget immune cell populations has not been fully elucidated. Moreover, the requirement for early intervention would preclude most patients from its benefits as over 80% of T1D subjects lack familial evidence and do not seek treatment until symptomatic when autoimmunity is well-developed,

thereby missing the critical window for early intervention. Thus, an approach that can be initiated within a wider window of time will be more reliable for T1D intervention, and a better understanding of both antigen-dependent and antigen-independent effects of DC therapy will assist in predicting the clinical outcome of DC therapy.

In T1D, T cell reactivity is initially limited to a few autoantigen determinants. However, as disease progresses, autoreactivity gradually expands intra- and intermolecularly to additional determinants and antigens, chronically recruiting naïve cells into the autoreactive pool and possibly leaving an altered immune repertoire with time, providing an explanation for why we observe the fall in efficacy of Ag-based therapies as the rise in autoimmunity expands [19–24]. This epitope spreading gives rise to an array of determinants that have distinct immunogenic properties and possibly unique roles in autoimmune pathogenicity. Regions within the whole antigen that T cells intrinsically recognize and respond to due to preferential antigen processing and presentation by antigen-presenting cells are known as dominant determinants (DD), while subdominant (SD) and ignored (ID) determinants are regions that are minimally unprocessed and unseen and fail to impact the naïve T cell repertoire. As autoreactivity expands to multiple determinants with time, it is expected that fewer T cells remain naïve to DD as they become recruited into a preprogrammed autoreactive response when challenged with a DD. In contrast, even in a late-stage disease, the naïve T cell pool should continue to remain nonreactive to SD or ID as they have had a minimal effect on the naïve T cell pool [25, 26]. Thus, DD-reactive T cells are progressively drained from the naïve pool, while uncommitted naïve T cells remain available to be potentially primed into regulatory function by SD and ID even at later stages of autoimmunity. Olcott et al. first examined this theory by treating NOD mice with a panel of control and T1D-specific autoantigen peptides during late-stage autoimmunity. They showed that only ID, but not target determinants (DD), could protect these mice from diabetes and that the ability of ID to prime Th2 responses did not attenuate with time [26].

In the present study, we hypothesized that through DC-guided presentation of SD or ID, we could better control the direction of the immune response to autoantigen challenge and quench established DD autoreactivity through regulatory T cell-biased bystander suppression. We investigated how various determinant peptides presented through immature DC therapy affected disease outcome when DC therapy was administered to NOD mice with active autoimmunity. In addition, we demonstrated antigen-independent effects of DC therapy and characterized changes in the overall immune response. The findings in this study will contribute to our current understanding on the role of antigen in DC-based therapies and guide the development of DC-based T1D immunotherapy.

2. Materials and Methods

2.1. Animals. Female NOD/ShiLtj (NOD), C57BL/6J (B6), and Balb/c mice were purchased from The Jackson Laboratory or Animal Care Services at the University of Florida. Bone marrow donor mice were 5–8 weeks of age. Up to five mice were housed together in micro isolator cages in a specific-pathogen-free (SPF) facility with access to food and water ad libitum. Mice were allowed to acclimate to the housing facility for one week prior to the initiation of any studies. Development of diabetes was monitored through twice weekly urine glucose testing using urine glucose test strips (Clinistix, Bayer). Upon detection of glucosuria, a small amount of blood was collected by pricking the tail vein and testing blood glucose using the Accuchek OneTouch glucose meter. A mouse with 2 consecutive daily readings of blood glucose greater than 250 mg/dl was considered to be diabetic. Mice were euthanized by CO_2 asphyxiation. All mouse experiments were performed in accordance with the University of Florida Institutional Animal Care and Use Committee.

2.2. Bone Marrow-Derived Dendritic Cells: Culture and Isolation. The femur and tibia were removed from mice and cleaned of muscle and connective tissue. The ends of the bones were cut, and bone marrow (BM) cells were flushed out with media using a 25–5/8 gauge needle attached to a syringe. Red blood cells were removed from bone marrow cells using ammonium chloride potassium (ACK) lysis buffer for 2 minutes at room temperature, then washed free of lysis buffer using PBS. BM-derived DCs were cultured in RPMI 1640 (Cellgro) supplemented with 10% fetal calf serum (FCS) (Invitrogen Life Sciences), 1x penicillin/streptomycin/neomycin (Gibco), and 10 mM HEPES buffer (Gibco) at a concentration of 10^6 cells/mL in flat-bottom 6-well culture plates (Corning). 500 U/mL GM-CSF (R&D Systems) and 1000 U/mL IL-4 (BD Pharmingen) were added to BM cultures to promote differentiation into DC. On day 2 or 3, half of the media was replaced with fresh media and cytokines. On day 5 or 6, cells were removed from the bottoms of wells with gentle pipetting and a cell scraper. DCs were purified using CD11c + positive selection magnetic beads (Miltenyi Biotec) and confirmed by flow cytometry to exceed 90% purity. Baseline expression of MHCII, CD80, and CD86 compared to DC stimulated for 24 h with TNF-α (semimature) or LPS (mature) was assessed by flow cytometry to characterize maturation state (Supplemental Figure 1).

2.3. Dendritic Cell Therapy. Dendritic cells for injection were derived from the bone marrow precursor cells of nondiabetic 4-–8-week-old female NOD mice. 100,000 DCs were suspended in 100 μl of sterile PBS for subcutaneous injection into the area of the hind footpads at 50 μl per footpad. Three weekly injections of PBS or peptide-pulsed or peptide-unpulsed DC (10^5 cells/mouse) were given to female NOD mice beginning at 9 weeks of age. Mice in short-term treatment studies were treated with one DC injection per week for three weeks, while mice in long-term treatment studies received the short-term treatment followed by boosters every other week. Boosters contained either 200 ng of corresponding peptide in PBS vehicle, or peptide-pulsed DC as received previously. Mice were monitored for normal locomotor

activity following footpad injections to ensure no disruption of accessibility to food and drink.

2.4. Flow Cytometry. Cells were prepared into single-cell suspensions in FACS buffer (1x PBS/1% FCS) and blocked in Fc Block CD16/32 (2.4G2). Antibody used to identify dendritic cells was CD11c (HL3). Antibodies used to characterize DC maturation were I-A^b [25-9-17], I-A^d (39-10-8, cross reacts with NOD I-A^{g7}), CD80 (16-10A1), and CD86 (GL1). Antibodies used to characterize T cells were CD3 (145-2C11), CD4 (RM4-5), and CD8a (53-6.7). Antibodies used to characterize B cells were B220 (RA3-6B2) and CD19 (1D3). We also used CD25 (PC61) and Foxp3 (FJK-16s) to assess regulatory T cell population, CD11b (M1/70) to assess macrophages, CD44 (IM7) and CD62 (MEL-14) to assess memory T cells, CD138 (281-2) for plasma cells, and CD80 (1610-A1) and CD35 (8C12) for memory B cells. Cells that were further examined for intracellular markers were fixed using Cytofix/CytoPerm reagent (eBioscience) for 15 minutes at room temperature, then washed in Perm/Wash (eBioscience). All subsequent steps were performed in Perm/Wash to maintain membrane permeability. Cells were analyzed by flow cytometer (FACS Calibur, BD Pharmingen). Live cells were gated from dead cells on the basis of forward/side scatter or with 7AAD (amino-antimycin D) labeling. Isotype controls include mouse IgG3κ, rat IgG2a, hamster IgG1κ, and hamster IgG1λ. All antibodies were purchased from BD Pharmingen or eBiosciences. FACS Calibur equipment (BD Biosciences) was used to collect flow cytometry data, and results were analyzed using FCS Express (De Novo).

2.5. Peptides. Peptides were purchased from Peptides International (Louisville, KY) and Bio-Synthesis Inc. (Lewisville, TX) and determined to be >90% purity by HPLC analysis. All peptides are tested to be endotoxin-free. Lyophilized peptides were dissolved in RPMI media at 1 mg/mL, then sterile filtered using a syringe apparatus (Gibco). Once resuspended in media, peptides were stored at 4°C as a working solution for up to 2 months. Lyophilized peptides were stored at −20°C indefinitely. Dominant determinants (DD) used were insulin β9-23 (SHLVEALYLVCGERG), and subdominant determinant (SD) used was $GAD65_{78\text{-}97}$ (KPCNCPKGDVNYAFLHATDL). Ignored determinant (ID) used was $GAD65_{260\text{-}279}$ (PEVK EKGMAALPRLIAFTSE).

2.6. Dendritic Cell Peptide Pulsing. DCs were pulsed with 3 μM of peptide in cRPMI for 1-2 h in a humidified incubator 37°C with 5% CO_2. Cells were washed 3 times and resuspended in PBS at 10^6 cells/mL for injection.

2.7. Proliferation Assay. Suspensions of spleen cells were in serum-free HL-1 media (Biowhittaker Cambrex) with the addition of penicillin/streptomycin/neomycin (Gibco) and L-glutamine (Gibco) in triplicate with a selected peptide (25 μM). Cells were cultured at 1×10^6 cells/well in round-bottom 96-well plates at 37°C. At 72 h of culture, 1 μCu 3H-thymidine (Amersham Biosciences) in 50 μl of media was added per well and allowed to incorporate for 12–16 h. Cells were harvested and washed using an automated cell harvester (Perkin Elmer), and radioactivity was analyzed using a liquid scintillation counter. cpm outliers identified by Grubbs test were removed from analysis.

In assessment of *in vitro* spontaneous proliferation of Tregs following DC therapy, CFSE-labeled spleen cells from female NOD mice from different groups were cultured in serum-free HL-1 media without stimulation and allowed to proliferate for 72–84 h. Cells were subject to surface staining for CD4 and subsequent intracellular staining for Foxp3 and analyzed for proliferating Foxp3+ cells on gated CD4+ T cells.

For assessment of whether IL-2, IL-7, or IL-15 was responsible for the *in vitro* spontaneous T cell proliferation of spleen cells from DC-treated mice, spleen cells were cultured at 1×10^6 cells/well in round-bottom 96-well plates at 37°C in serum-free HL-1 media without stimulation in the presence of isotype IgG antibody, or neutralizing anti-IL-2 antibody, anti-IL-7 antibody, or anti-IL-15 antibody for 84 h. Thereafter, 1 μCu 3H-thymidine (Amersham Biosciences) in 50 μl of media was added per well and allowed to incorporate for 12–16 h. Cells were harvested and washed using an automated cell harvester (Perkin Elmer), and radioactivity was analyzed using a liquid scintillation counter.

For evaluating homeostatic proliferation in normal and autoimmune mouse models, NOD, B6 mice were treated with 3 weekly subcutaneous injections of DC (10^5/injection) or PBS beginning at 9 weeks of age, and Balb/c mice at the same age were treated with 3 weekly intravenous injections of DC or PBS. Spleen cells were prepared 2 weeks following final injection to assess 3H-thymidine proliferation in the HL-1 media in the absence of *in vitro* stimulation.

2.8. Suppressor Assay. Spleen cells were prepared and suspended in MACS buffer. CD4+ cells were enriched through depletion of unwanted cells using the CD4+CD25+ Regulatory Cell Isolation Kit (Miltenyi Biotec). Next, CD25 + cells were positively selected from the preenriched fraction following the instruction from the manufacturer (Miltenyi). Suppressor CD4+CD25+ cells were cultured with CD4 +CD25+ depleted cells (10^5) at 0 : 1, 1 : 2, and 1 : 4 ratios in a round-bottom 96-well plate. Cells were cultured in serum-free HL-1 media with anti-CD3e (0.05 μg/200 μl well). At 72 h of culture, 1 μCu 3H-thymidine (Amersham Biosciences) in 50 μl of media was added per well and allowed to incorporate for 12–16 h. Cells were harvested and washed using an automated cell harvester (Perkin Elmer), and radioactivity was analyzed using a liquid scintillation counter. The suppression rate = (proliferation (cpm) without CD4+CD25+ T cells − proliferation (cpm) with CD4+CD25 + T cells)/proliferation (cpm) without CD4+CD25+ T cells.

2.9. ELISA for Global Suppression Analysis. Eight-week-old female NOD mice received PBS, unpulsed, or peptide-pulsed DC injections as described previously, once weekly for three consecutive weeks. One week following the last injection, mice were immunized in the footpad with 100 μg/mouse of Keyhole limpet hemocyanin (KLH)

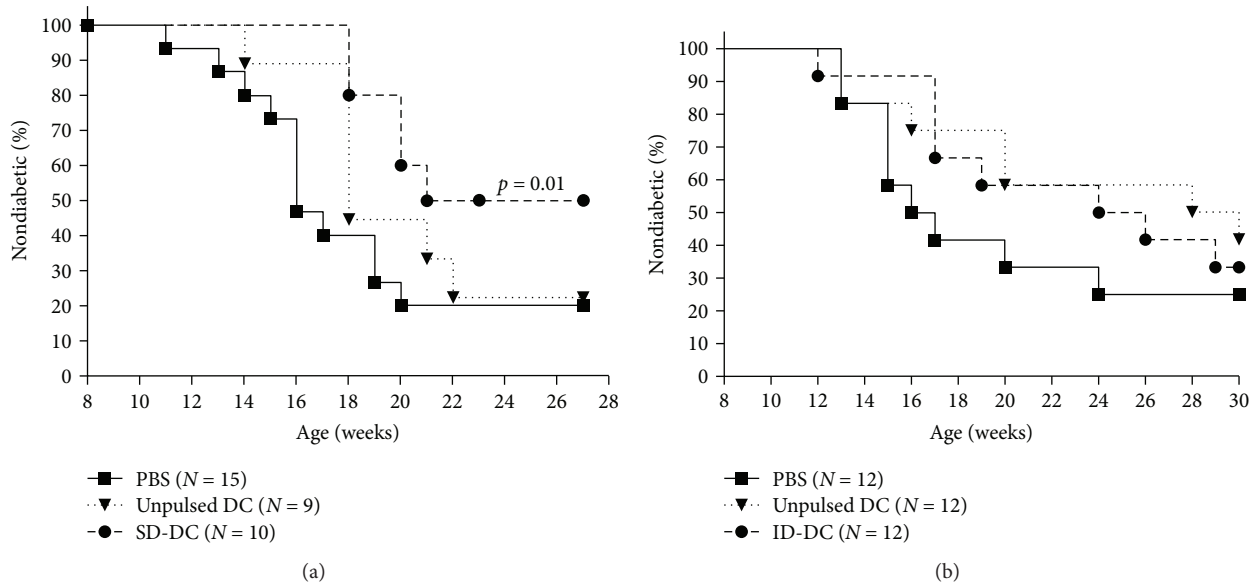


Figure 1: Injections of immature DCs pulsed with subdominant and ignored dominant β cell antigenic peptides significantly delay T1D in NOD mice. (a) Nine-week-old NOD mice received subcutaneous injection of PBS, unpulsed DCs, or subdominant determinant-pulsed DCs, once a week for 3 weeks. Then, the mice were monitored for diabetes onset till 27 weeks of age. (b) Nine-week-old NOD mice received subcutaneous injection of PBS, unpulsed DCs, or ignored determinant peptide-pulsed DCs, once a week for 3 weeks. Then, the mice were monitored for diabetes onset till 30 weeks of age. Kaplan–Meier survival curves were depicted, and statistical analysis was performed using Log-Rank test; $p < 0.05$ is considered statistically significant.

(Calbiochem) in Alum (Pierce) weekly for two weeks. Ten to fourteen days following the final KLH immunization, serum samples were collected from mice for the detection of antibodies to KLH by ELISA (Life Diagnostics).

2.10. BrdU Incorporation to Assess In Vivo Immune Cell Homeostatic Proliferation. Mice received daily intraperitoneal injections of BrdU (bromodeoxyuridine) in sterile PBS (2 mg/100 μl/mouse) for 4 days, then sacrificed 1-2 days following final injection to tissue for analysis of BrdU incorporation. Spleens, livers, and pancreata were fixed in 10% formalin at room temperature for 24–48 hours. Tissues were embedded in paraffin and sectioned at 4 μm for staining using anti-BrdU-HRP Ab and DAB detection and counterstained with hematoxylin. Two sections per sample were collected 100 micron apart for analysis using Aperio's Spectrum ScanScope imaging software. The frequency of BrdU-positive cells was determined using ScanScope's image analysis algorithm that detects positively stained cells on the basis of programmed color and saturation sensitizers within a measured tissue area. Percent BrdU positive is calculated as area positive/area total.

2.11. Statistical Analysis. Data were analyzed using the Kaplan–Meier survival curve with Gehan-Breslow-Wilcoxon test to determine if treatment provided protection. Student's *t*-test was also used to identify statistical differences. The Grubbs' test identifies outliers in triplicate wells of proliferation assays. A criterion of $p < 0.05$ was used to define significance.

3. Results

3.1. Bypassing of Natural Antigen Processing Using DC Pulsed with Underpresented Autoantigen Peptides Leads to T1D Protection in NOD Mice with Active Autoimmunity. Antigen-based studies in mice have demonstrated that DD are ineffective for tolerance induction when applied as peptide therapy in NOD mice with progressive insulitis, and emerging data suggest that use of nontargeted determinants may allow better priming of naïve T cells into regulatory function if treatment is initiated after the autoimmune process is well-established [26]. While SD and ID determinants may be able to better prime regulatory responses from naïve T cells, their reduced or lack of constitutive presentation may require lifelong treatment to maintain the regulatory T cell pool. Thus, we first aimed to assess whether short-term DC therapy pulsed with subdominant determinants better protected NOD from T1D compared to unpulsed DC. We treated 9-week-old female NOD mice with three treatments of bone marrow-derived immature DC unpulsed, or pulsed with synthetic peptides of SD. As shown in Figure 1(a), we found that only recipients of SD-DC, but not PBS or unpulsed DC, were protected from T1D ($p = 0.01$). Of note, SD-DC were able to significantly delay T1D in 100% of SD-DC recipients through the 17th week of age while 40% of PBS controls became diabetic. This suggests that complete protection was conferred for over 8 weeks, and the protection was not durable for the life of the animal. However, complete protection would be ideal in the clinical setting. As ID do not naturally elicit T cell responses, we hypothesized that a larger pool of naïve T cells responding to ID would remain available for priming into

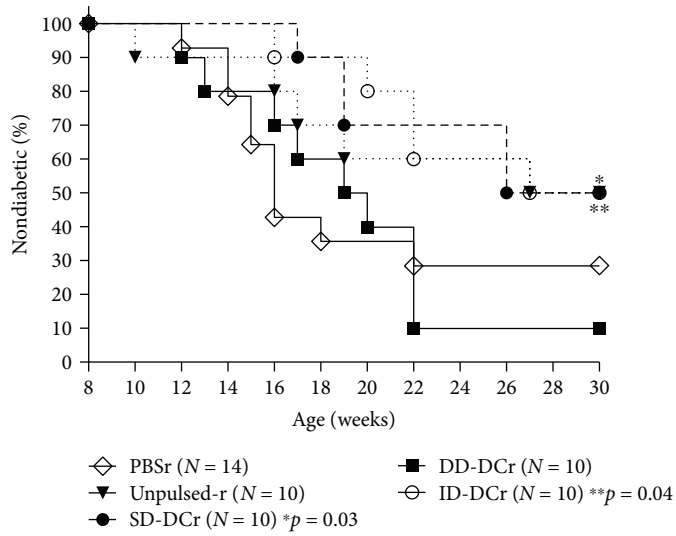


Figure 2: DC therapy-induced T1D protection can be maintained by ignored or subdominant determinant antigenic peptide boosters. Nine-week-old NOD mice received three weekly injections of PBS, unpulsed DCs, or DCs pulsed with DD, SD, or ID peptides. Thereafter, the mice received the corresponding treatment every other week until the study ended. Diabetes onset was monitored once a week. *P* values represent difference compared between PBS and treatment groups.

tolerance compared to SD. This advantage in available naïve T cell pool size may translate into better protection. Therefore, we performed another study using ID-pulsed DC in 9-week-old NOD mice with active autoimmunity. We administered three weekly injections of PBS, unpulsed, or ID-pulsed DC to mice and observed them for the development of T1D. Surprisingly, we found that ID-DC treatment was not able to significantly protect mice from T1D though we did observe an initial delay in T1D development (Figure 1(b)).

3.2. Repetitive Administration of DC Pulsed with ID or SD Prolongs T1D Protection. To assess whether the lack of constitutive presentation of the ID accounted for the loss, we refined this study to include repetitive injections that allowed for consistent presentation of the normally unpresented determinants. Since cell procurement in the clinical setting is both costly and labor-intensive, we wanted to first elucidate whether peptide-only boosters following short-term peptide-pulsed DC could maintain the protection. Because the fate of peptide therapy in the absence of a DC carrier is unknown, in a separate group of mice, we also followed the initial short-term priming treatment with peptide-pulsed DC boosters as proof of principle to account for any peptide competition that may occur *in vivo*. Boosters were given every other week until the end of the study. We found that peptide-only boosters could not continue protection (data not shown). However, as shown in Figure 2, repetitive SD ($p = 0.01$) or ID-pulsed DC treatment was protective ($p = 0.03$) in contrast to PBS control group. No protection was observed in mice receiving repetitive PBS or DD-pulsed DC treatment.

3.3. T1D-Specific Peptide-Pulsed DC Therapy Does Not Alter Immune Response to Non-T1D Antigen Challenge in Terms of Development of Antigen-Specific Antibodies. Because we observed an initial delay in development of T1D in all mice receiving DC therapy, we were uncertain whether the apparent DC-induced protection against T1D was actually due to an overall dampening of the immune response. We sought to evaluate whether DC therapy conferred specific protection against T1D, or whether the observed protection was an artifact of global immunosuppression that renders mice tolerant to all immune challenges. We tested this by evaluating the ability of DC-treated mice to respond to a non-T1D-specific antigen challenge. We administered either PBS, unpulsed, or ID-pulsed DC therapy as described previously, then immunized the mice with keyhole limpet hemocyanin (KLH), a protein commonly used to examine and elicit immune responses. Two weeks following KLH immunization, we collected sera from the treated mice to detect if an antibody response was mounted against KLH. As shown in Figure 3, there was no difference in the ability of DC-treated mice to generate an antibody response to KLH challenge as compared to PBS-treated mice ($p > 0.05$), suggesting that normal immune processes were intact and the protection previously observed can be attributed to T1D-specific protection.

3.4. Homeostatic Lymphocyte Proliferation Is Observed following DC Therapy: Immediate and Sustained Effects. In our studies, we observed that antigen pulsing with SD or ID determinants improved disease outcome. However, mice receiving unpulsed DC also seemed to exhibit a delay in T1D development though they did not achieve significant protection. Since protection from unpulsed DC therapy has been reported in early intervention studies, we wanted to assess how DC therapy affected the immune response as a whole including antigen-independent responses. The spleen is a major site of immune cell interactions and antigen processing, with active processes that contribute to the overall immune status [27, 28]. Thus, we sought to examine

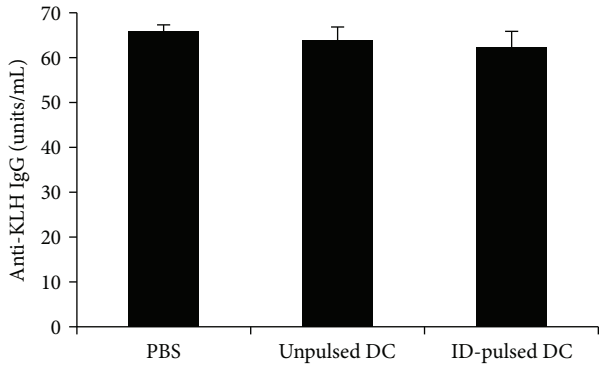


Figure 3: Antibody response following KLH immunization in control and DC-treated mice. NOD mice were treated with 3 weekly injections of either PBS or unpulsed DCs or ID-pulsed DCs ($N = 3$/group). Two weeks following DC therapy, mice were immunized with 2 weekly injections of KLH. Serum antibody levels were assessed 14 days following final KLH immunization. The levels of anti-KLH antibodies of each group are shown as mean ± SD.

cellular responses in this immune cell-rich environment. To evaluate the spleen cell response following DC therapy, we cultured spleen cells with and without autoantigen peptide stimulation for 86 hours, then observed for proliferation using 3H-thymidine incorporation. We found that even in the absence of *in vitro* peptide stimulation, spleen cells isolated from all DC-treated mice had 3–14-fold increase in proliferation compared to PBS-treated mice (Figure 4). This effect of spontaneous proliferation was enhanced in mice receiving antigen-pulsed DC but did not increase with recall peptide challenge suggesting that the response was not eliciting a pathogenic reactivity to the immunizing peptide. The proliferation was seen as soon as just 2 weeks following the last DC treatment at 14-week age (Figure 4(a)) and continued into 40 weeks of age, 29 weeks after the cessation of treatment (Figure 4(b)).

3.5. Homeostatic Proliferation Occurs in Healthy and Autoimmune Mouse Strains following DC Therapy. Homeostatic proliferation has been linked to immunodeficiency which promotes a compensatory expansion of "immunological space" [29]. Because NOD mice have been shown to have abnormalities in the immune function of many cell types including differences in DC phenotype and function [30–34], we evaluated whether this homeostatic proliferation was a true effect of DC therapy or only an effect associated with immunotherapy in an animal afflicted with aberrant immune cell subsets. We administered DC therapy to the autoimmune NOD mouse model as well as the healthy control mouse models C57BL/6J and Balb/c and evaluated spleen cell proliferation. As depicted in Figure 4(c), spleen cell homeostatic proliferation following DC treatment occurred in both NOD and nonautoimmune-prone mouse models, suggesting that DC therapy uniquely resulted in a reprogramming of immune cell homeostasis. Additionally, this pattern was independent of route of administration, as Balb/c mice were treated with intravenous DC injections while NOD and B6 mice were given subcutaneous injections.

3.6. Homeostatic Proliferation Is Driven by Interleukin-2. Our experiments comparing NOD mice to healthy control C57BL/6J and Balb/c mice revealed that spontaneous proliferation following DC treatment is not attributed to lymphopenia possibly happening in NOD mice. Flow cytometric phenotyping of the proliferating cells did not provide evidence for CD4+CD44^{hi}CD62^{lo} memory T cell nor CD80 +CD35+ memory B cell expansion (data not shown). Another mechanism driving the expansion may be soluble cytokines that contribute to proliferation or maintenance of homeostasis. Studies have shown that IL-2 and IL-15 can activate NK, T, and B cells, induce their proliferation and survival, and stimulate cytokine production [35, 36]. IL-7, a related cytokine sharing the common gamma chain, has been shown to have a role in T cell development, homeostatic proliferation, and survival [35, 37, 38]. Thus, we performed proliferation assays in the presence of cytokine neutralizing antibodies to assess whether proliferation could be abated. We found that neutralization of IL-7 or IL-15 had a minor effect on cell proliferation, while culture with IL-2 neutralizing antibody significantly reduced the expansion of spleen cells of DC-treated mice by 45%, twice the effect observed from cells of PBS-treated mice (Figure 5).

3.7. DC Therapy Results in Sustained Expansion of Regulatory T Cells with Enhanced Immunosuppressive Function. Our work has shown that DC therapy protects mice from T1D and induces noninflammatory homeostatic proliferation of CD4+ T cells. Evidence from the literature suggests that a possible mechanism for protection from DC therapy is the induction of regulatory T cells. Thus, we sought to examine whether Tregs are being induced and whether they are part of the proliferating cell population. Following unpulsed and antigen-pulsed DC therapy, we looked for changes in regulatory T cell frequency and function by evaluating the proportion of CD4+Foxp3+ cells in DC-treated and control mice and examining their ability to suppress proliferation of effector cells. As shown in Figure 6(a), we found that there was an over 2-fold increase in the frequency of CD4+Foxp3+ T cells in mice receiving unpulsed DC and an over 4-fold increase in frequency of CD4+Foxp3+ T cells in mice receiving ID DC, demonstrating that DC therapy resulted in sustained expansion of regulatory T cells and that the effect was particularly enhanced in mice receiving ID-pulsed DC. This homeostatic expansion of Tregs was independent of *in vitro* peptide stimulation, as the pattern was observed in both stimulated (data not shown) and unstimulated cell cultures.

We also examined whether there were functional differences in regulatory T cells following DC therapy. We performed a suppressor cell function assay by coculturing CD4+CD25+-depleted spleen cells with CD4+CD25+-purified cells at ratios of 0 : 1, 1 : 2, and 1 : 4 in the presence of anti-CD3. As seen in Figure 6(b), regulatory T cells from both unpulsed and peptide-pulsed DC-treated mice demonstrated greater suppressive function in a dose-dependent manner, with the effect enhanced in the peptide-pulsed DC group. The enhanced suppression was found to be nearly 2-3-fold

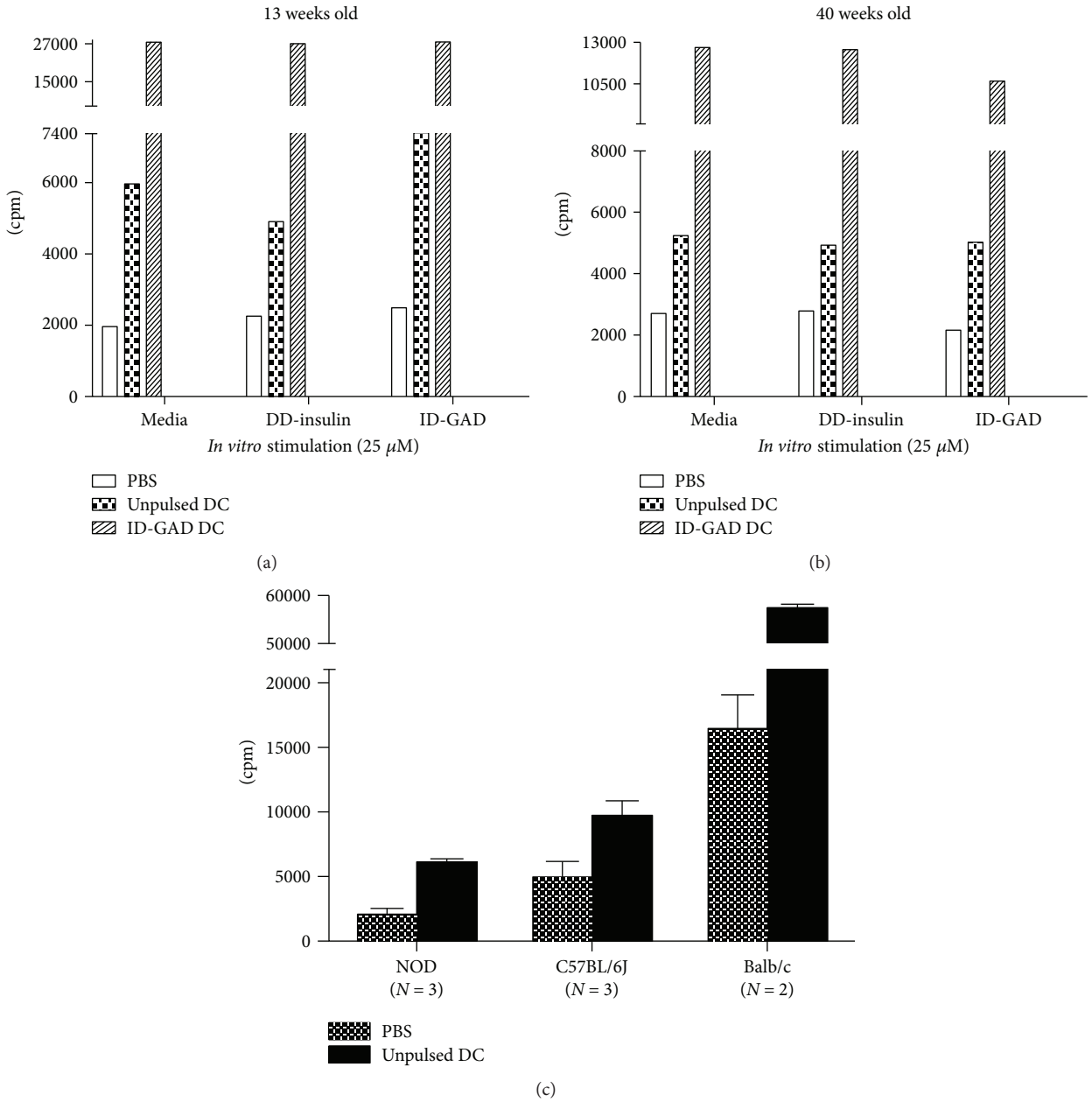


Figure 4: Spleen cell homeostatic proliferation following DC therapy. (a) Nine-week-old NOD mice received subcutaneous injection of PBS, unpulsed DCs, or ID peptide-pulsed DCs, once a week for 3 weeks. Spleen cells from NOD mice of each group at 13 weeks of age were cultured in serum-free HL-1 media alone or with DD-insulin, or ID-GAD for 86 h, and 3H-thymidine was added for incorporation during the final 16 h of culture. Proliferation was assessed by liquid scintillation quantification of counts per minute (cpm). Data shown are the mean cpm (counts per minute) of triplicate values from one of ten experiments. (b) Nine-week-old NOD mice received subcutaneous injection of PBS, unpulsed DCs, or ID peptide-pulsed DCs, once a week for 3 weeks. Spleen cells from NOD mice of each group at 40 weeks of age were cultured in serum-free HL-1 media alone or with DD-insulin, or ID-GAD for 86 h, and 3H-thymidine was added for incorporation during the final 16 h of culture. Proliferation was assessed by liquid scintillation quantification of counts per minute (cpm). Data shown are the mean cpm (counts per minute) of triplicate values from one of ten experiments. (c) Homeostatic proliferation was observed in healthy and autoimmune mouse models. NOD, B6 mice were treated with 3 weekly subcutaneous injections of DC (10^5/injection) or PBS beginning at 9 weeks of age, and Balb/c mice were treated with intravenous injection of DC or PBS at the same age. Spleen cells were collected 2 weeks following final injection to assess 3H-thymidine proliferation in the HL-1 media in the absence of *in vitro* stimulation. Data shown are the mean cpm (counts per minute) ±SD.

greater in DC-treated mice at a 1 : 2 ratio. This effect was magnified when the ratio was decreased to 1 : 4, where up to a 10-fold enhancement in suppression was observed. These results demonstrate that on a cell-to-cell level, regulatory T cells isolated from DC-treated mice are more potent in suppressor function than those isolated from PBS-treated mice.

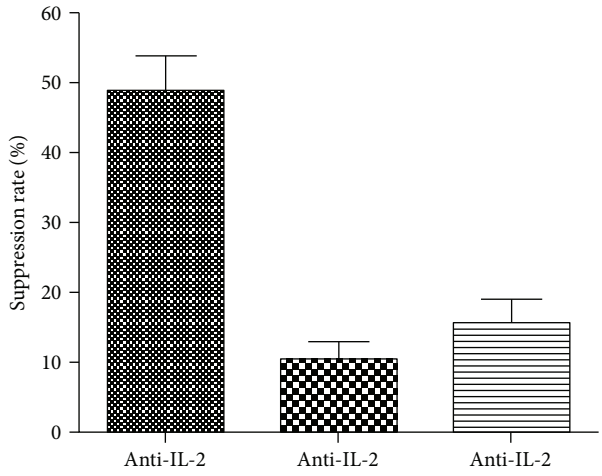


Figure 5: IL-2 is a contributing cytokine for the proliferating cell populations. Spleen cells prepared from different groups shown in the figure were cultured in serum-free HL-1 media for 86 hours in the presence of isotype control Ab, or neutralizing Ab against IL-2, IL-7, or IL-15. 3H-thymidine was added to culture for the last 16 hours. Data shown is representative of 10+ experiments collected through a range of posttreatment time points (13–41 weeks of age).

4. Discussion

T1D is a dynamic autoimmune disorder characterized by T cell-mediated destruction of pancreatic islets driven by an expanding T cell autoreactivity toward β cell autoantigens. Dendritic cells, which present antigen and direct T cell responses, are an ideal platform for use in T1D treatment as DC therapy could potentially correct the specific underlying autoimmune aberrancy in T1D. DC therapy can uniquely control (1) the direction of the immune response through the selection of either immunogenic or tolerogenic classes of DC, as well as (2) dictate the target antigen that the response is directed toward through the presentation of a chosen antigen, reinforcing DC therapy to be an effective and powerful strategy for immune modulation. Reports of DC therapy for tolerance induction have been successfully demonstrated when applied before or in the early stages of autoreactivity in animal models of various autoimmune diseases, as well as in studies of transplant/graft acceptance [39–45]. However, if treatment is initiated after the autoimmune process is active, efficacy in DC-mediated protection declines. While NOD mice have a predictable timeline for T1D onset allowing for intervention to be planned accordingly, the dynamics of autoreactivity processes in human has been difficult to define due to multiple variations in subtypes that compound assessment. Additionally, the majority of subjects susceptible to T1D lack familial history that would otherwise prompt early autoantibody screening; thus, the opportunity for early intervention in humans is low, emphasizing the need for therapy that can treat both established and new onset disease.

We sought to understand how to better develop DC therapy for translation into the clinical setting. To create DC for therapy with more durable protection, we considered another aspect of DC therapy: selection of antigen for loading prior to infusion. We and others have demonstrated that the administration of β cell autoantigens in a tolerogenic modality is highly effective in preventing T1D in the NOD mouse [1, 2, 22, 46–49]. However, uncertainties in extrapolating appropriate antigen doses and correlating treatment timeline have hindered its translation into the clinical setting, particularly since studies have shown that the immune response can pivot toward immunity or tolerance depending on antigen dose. Fortunately, antigen presentation in the context of a tolerogenic DC may circumvent the issue of ambiguous immune deviation associated with antigen treatment alone. Based on work from Kaufman's group, we believed that dominant determinants (DD) identified to be the initiators of the autoimmune response chronically recruit naïve T cells into the pathogenic pool; thus, the readministration of these determinants only reactivated cells that were programmed to respond pathogenically [26]. However, subdominant determinants (SD) or ignored determinants (ID), which have a minimal impact on naïve T cell activation, should have large pools of naïve T cells available for priming into tolerance when we bypass natural antigen processing to experimentally present these peptides. We compared the efficacy of DD, SD, and ID peptide classes in DC therapy to protect 9-week-old NOD mice and found that only SD- and ID-pulsed DC were able to protect mice when the treatment was applied in NOD with active ongoing autoimmunity (Figure 2). Specifically, just three weekly injections of 1×10^5 SD-DC protected NOD from T1D with a significant delay in the onset of T1D, though complete protection was not achieved. We examined whether ID-DC, which should have a comparatively larger pool of naïve T cells to prime into tolerance, would be more effective in conferring protection. However, we found that three injections of ID-DC were not sufficient to achieve prolonged protection, as a sudden increase in diabetes onset within 6 weeks of the last treatment dampened the treatment success. We speculated that since ID are not constitutively presented, treatment may need to be continued to maintain the regulatory T cell pool. We treated another cohort of mice as previously described, then followed by a series of priming injections with boosters every other week. We found that mice receiving boosters of SD-DC or ID-DC were significantly protected from T1D. This enhanced protection was not seen in mice treated with repetitive injections of PBS or unpulsed or DD-pulsed DC, suggesting that the protection is attributed to the nature of the antigens. Because cell procurement is a labor-intensive and costly treatment, we were also interested in determining whether peptide-only boosters following the initial priming series could effectively maintain the same protection. Unfortunately, mice receiving peptide-only boosters following the initial DC priming developed T1D with age (data not shown). It is possible that our selected peptide-boosting dose was not optimal to maintain tolerance, or the peptide presented by the host antigen-presenting cells altered T cell functionality.

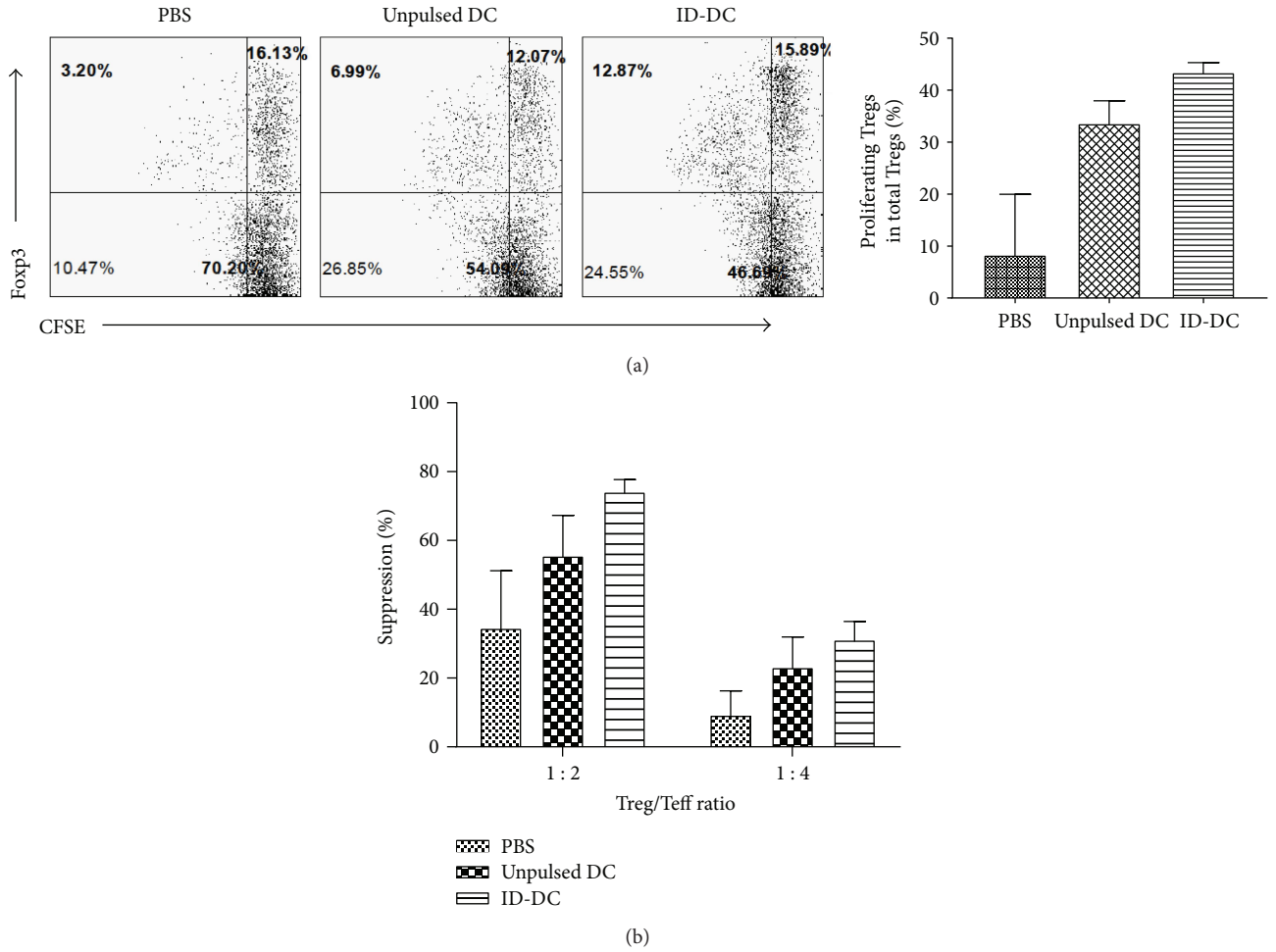


Figure 6: Assessment of Treg spontaneous proliferation and function induced by DC therapy. (a) Assessment of *in vitro* spontaneous proliferation of Tregs following DC therapy. CFSE-labeled spleen cells from female NOD mice from different groups were cultured in serum-free media without stimulation and allowed to proliferate for 72–84 h. Cells stained with CD4 and Foxp and analyzed by flow cytometry. The proliferating Foxp3+ cells were analyzed by gating on total CD4+ cells. Data shown is representative of 3 experiments from mice aging from 13–41 weeks old. (b) For suppressor T cell assay. Female 9-week-old NOD mice were treated with 3 weekly injections of DC, then Treg function was assessed at 13 weeks of age. CD4+CD25+ Tregs were purified and cocultured with CD4+CD25+ T cell-depleted spleen cells at ratios of 0 : 1, 1 : 2, and 1 : 4 and stimulated with anti-CD3 antibodies (0.05 μg/ml). Proliferation was assessed by 3H-thymidine incorporation. The suppression rate = (proliferation (cpm) without CD4+CD25+ T cells − proliferation (cpm) with CD4 +CD25+ T cells)/proliferation (cpm) without CD4+CD25+ T cells.

Alternatively, this finding stresses the importance of the role of DC in therapy. Administration of peptide alone to a host with an existing aberrant immune system may be futile. Likewise, it is unknown what happens to peptide without using DC as an antigen carrier because peptide competition *in vivo* could render the injected peptide irrelevant. Overall, our findings indicate that antigen presentation, and particularly the class of determinant, plays an important role in DC-based immune modulation.

Consistent with this finding, we observed that antigen-DC-treated mice had a greater number of pancreatic islets compared to unpulsed-DC-treated mice (data not shown). We failed to observe β regeneration in all groups. Thus, it is possible that the islet preservation we observed was achieved through the induction of regulatory T cells that was enhanced with antigen-pulsed DC treatment, but this treatment was not enough to completely quench the inflammation generated from the pathogenic T cells.

To exclude the possibility that the observed protection from T1D was due to global immunosuppression, we examined whether NOD mice could generate a normal immune response to a non-T1D-related antigen challenge following treatment with DC therapy. We immunized PBS, unpulsed DC, and ID-DC-treated mice with KLH and examined their serum antibody responses. All DC-treated mice were able to mount antibody responses to KLH in a manner comparable to PBS controls, suggesting that normal immune responses were intact and the previously observed protection could be attributed to diabetes-specific protection.

Much of the current knowledge on how DC therapy affects immune responses has been delineated from studies with a focus on antigen-specific immune modulation, as we

have shown that antigen-based DC therapy-mediated protection is limited to the suppression of autoreactive processes specific to T1D. However, whether DC therapy results in antigen-nonspecific immune changes has not been well investigated. Recent evidence suggests that β cell antigens in T1D immunotherapy might not be necessary for therapy-induced protection [50]. The spleen is a major site of immune cell interactions and antigen processing, with active processes that contribute to the overall immune status [27, 28]. Thus, we sought to examine the spleen cell response in the absence and presence of T1D peptide stimulation. To our surprise, we found robust *in vitro* homeostatic proliferation of spleen cells isolated from DC-treated mice (unpulsed or antigen-pulsed), but not PBS-treated mice. The resulting changes are not antigen specific as we find this reprogrammed spleen cell responses in the absence of antigen stimulation. Furthermore, this effect was immediate and sustained, as the proliferation could be observed as early as just two weeks following only 2 DC injections, and was durable even 29 weeks after treatment had ended. Further characterization of these cells revealed that the proliferation was predominantly attributed to B and T lymphocytes (data not shown). In addition, by screening T cell proliferation-related cytokines in this DC therapy-induced spontaneous proliferation of lymphocytes, IL-2 was found to be the responsible cytokine while IL-7 and IL-15 played a minor role.

Because NOD mice have a defect in DC phenotype and function, we evaluated whether this homeostatic proliferation was a true effect of DC therapy or an only effect associated with therapy using DC with an aberrant phenotype. We treated the nonautoimmune-prone mouse strains Balb/c and C57BL/6J mice with PBS or DC and observed a similar enhancement in homeostatic proliferation in mice receiving DC therapy, confirming that the effect is a true immune response to DC therapy.

To determine whether homeostatic proliferation occurred *in vivo*, we treated mice with BrdU and collected spleens to detect for BrdU incorporation. We were not able to detect a difference in percentage of proliferating cells between mice treated with PBS compared to DC-treated mice (data not shown). We also sought to determine whether the protection observed resulted from increased beta cell regeneration. We examined pancreata and found no differences in BrdU incorporation.

In T1D, the lack of an adequate regulatory response allows autoreactive T cells to become pathogenic, thereby invading and destroying the pancreatic islet cells. Multiple studies have demonstrated that DC therapy can confer protection against autoimmunity through the induction of regulatory T cells that inhibit the pathogenic T cell inflammation [44, 51–64]. Thus, we evaluated the effect of DC therapy on the regulatory T cells. Consistent with our earlier finding of homeostatic CD4+ T cell expansion, we found that DC therapy resulted in a durable 2-3-fold increase in the frequency of proliferating CD4+Foxp3+ regulatory T cells when cultured *in vitro*, and this effect was further enhanced using ID-DC. DC therapy also enhanced the immunosuppressive function of Tregs, as CD4+CD25+ regulatory cells from DC-treated mice more potently suppress anti-CD3 antibody-stimulated proliferation of CD4+CD25-depleted responder cells compared to Tregs from PBS controls. These results suggest that DC (esp. antigen-pulsed) therapy primes generation of more immunosuppressive Tregs. Again, the addition of antigen to DC therapy leads to even greater enhanced suppressive function. Of interest, these findings were observed in both nondiabetic and delayed diabetic mice emphasizing the correlation to DC treatment. While we observed an increase in both frequency and function of Tregs with DC treatment *in vitro*, we did not observe a correlation in protection from T1D *in vivo*. A potential explanation was proposed by work from Diane Mathis's group, which demonstrates that while defects in NOD Tregs contribute to T1D, it may be an effect of overresponsive effector T cells to self-antigen that truly drive the immunopathology [65]. Thus, the loss of tolerance may be related not to impaired function or decreased frequency of NOD Tregs, but rather a decline in the ability of NOD T cell effectors to respond to fully competent Tregs. However, our studies of Treg function examine PBS-treated Tregs versus DC-treated Tregs against NOD effectors, which in concept should be similarly impaired, so our observation of functional differences between the treatment groups can be attributed to a true variation between PBS- and DC-treated Tregs. Alternatively, it is possible that the improvement was not sufficient to pivot the balance in favor of regulation in the presence of a potent inflammatory effector T cell response that has been shown to grow with age [65]. This may be supported by our observation of increased islet survival in antigen-pulsed-DC-treated mice, but not unpulsed-DC-treated mice that have similar levels of lymphocyte infiltrate (data not shown), as the undefined lymphocyte population may potentially be an influx of both pathogenic and regulatory T cells. The influence of regulatory T cells to preserve existing islets must surpass the destruction by pathogenic T cells to maintain physiologically relevant numbers of functional islets for metabolic control of insulin. Nonetheless, the later stages of advanced autoimmunity immediately prior to T1D onset may simply not be amenable to a one-armed intervention; Tregs alone may not be sufficient to rescue β cell death and the requirement for combinatorial strategies to treat both autoimmunity and regenerate β cell mass may become necessary. To assess the T1D-protective regulatory T cells, an alternative approach to be considered is to use adoptive T cell transfer to examine whether the DC-treatment-induced regulatory T cells are more potent in protecting NOD mice from T1D or protecting NOD-Rag−/− mice from diabetes induced by cotransferred diabetogenic T cells, compared to regulatory T cells from PBS-treated mice.

Collectively, our study demonstrates that DC therapy results in antigen-dependent and antigen-independent effects on immune modulation [66]. We find that the selection of autoantigen peptide for therapies aimed to prime naïve T cells has a critical impact on the efficacy of protection against dynamic autoimmune diseases. Constitutively, underpresented autoantigen determinants may be more effective in tolerance induction when used in late-stage intervention. This fundamental principle of altering

native determinant presentation to accommodate a changing T cell repertoire can be extended to the design of treatment for any dynamic autoantigen-based diseases. We also demonstrate that immature DC therapy augments the immune response in an antigen-independent manner resulting in homeostatic expansion of functionally enhanced Tregs. Overall, these findings demonstrate the durable potency of DC therapy in the modulation of antigen-specific and antigen-nonspecific immune responses and provide an important step toward translation into the clinic as other peptide-based therapies for T1D have been limited to early intervention.

Disclosure

This paper represents part of the doctoral dissertation of Jeannette Lo.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Authors' Contributions

Jeannette Lo performed the experiments and wrote the manuscript; Chang-Qing Xia performed part of the experiments and wrote the manuscript; Ruihua Peng performed some of the experiments; Michael J. Clare-Salzler supervised the study and edited the manuscript; Jeannette Lo and Chang-Qing Xia share first authorship.

Acknowledgments

This work was supported by NIH/NIDDK/NIAID R21/R33 to Michael J. Clare-Salzler and partially supported by NIH1R21DK08026-01A2 to Chang-Qing Xia.

Supplementary Materials

Figure 1: characterization of DC phenotype. DCs were differentiated in the presence of GM-CSF and IL-4 for 5-6 days (immature iDC), and baseline expression of MHCII, CD80, and CD86 was assessed by flow cytometry compared with DC stimulated with TNFα (semimature smDC) or LPS (mature mDC) for 24 h. *(Supplementary Materials)*

References

[1] J. Tian, M. A. Atkinson, M. Clare-Salzler et al., "Nasal administration of glutamate decarboxylase (GAD65) peptides induces Th2 responses and prevents murine insulin-dependent diabetes," *The Journal of Experimental Medicine*, vol. 183, no. 4, pp. 1561–1567, 1996.

[2] D. Daniel and D. R. Wegmann, "Protection of nonobese diabetic mice from diabetes by intranasal or subcutaneous administration of insulin peptide B-(9-23)," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 2, pp. 956–960, 1996.

[3] D. Elias, T. Reshef, O. S. Birk, R. van der Zee, M. D. Walker, and I. R. Cohen, "Vaccination against autoimmune mouse diabetes with a T-cell epitope of the human 65-kDa heat shock protein," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 8, pp. 3088–3091, 1991.

[4] J. M. Barker, K. K. McFann, and T. Orban, "Effect of oral insulin on insulin autoantibody levels in the diabetes prevention trial type 1 oral insulin study," *Diabetologia*, vol. 50, no. 8, pp. 1603–1606, 2007.

[5] P. Pozzilli, "The DPT-1 trial: a negative result with lessons for future type 1 diabetes prevention," *Diabetes/Metabolism Research and Reviews*, vol. 18, no. 4, pp. 257–259, 2002.

[6] J. Ludvigsson, "Therapy with GAD in diabetes," *Diabetes/Metabolism Research and Reviews*, vol. 25, no. 4, pp. 307–315, 2009.

[7] C. D. Agardh, K. F. Lynch, M. Palmer, K. Link, and A. Lernmark, "GAD65 vaccination: 5 years of follow-up in a randomised dose-escalating study in adult-onset autoimmune diabetes," *Diabetologia*, vol. 52, no. 7, pp. 1363–1368, 2009.

[8] I. Raz, A. Avron, M. Tamir et al., "Treatment of new-onset type 1 diabetes with peptide DiaPep277® is safe and associated with preserved beta-cell function: extension of a randomized, double-blind, phase II trial," *Diabetes/Metabolism Research and Reviews*, vol. 23, no. 4, pp. 292–298, 2007.

[9] M. Feili-Hariri and P. A. Morel, "Phenotypic and functional characteristics of BM-derived DC from NOD and non-diabetes-prone strains," *Clinical Immunology*, vol. 98, no. 1, pp. 133–142, 2001.

[10] X. Chen, L. H. C. Makala, Y. Jin et al., "Type 1 diabetes patients have significantly lower frequency of plasmacytoid dendritic cells in the peripheral blood," *Clinical Immunology*, vol. 129, no. 3, pp. 413–418, 2008.

[11] A. C. Vasquez, M. Feili-Hariri, R. J. Tan, and P. A. Morel, "Qualitative and quantitative abnormalities in splenic dendritic cell populations in NOD mice," *Clinical and Experimental Immunology*, vol. 135, no. 2, pp. 209–218, 2004.

[12] P. Chaturvedi, B. Agrawal, M. Zechel, E. Lee-Chan, and B. Singh, "A self MHC class II β-chain peptide prevents diabetes in nonobese diabetic mice," *The Journal of Immunology*, vol. 164, no. 12, pp. 6610–6620, 2000.

[13] J. Lo, R. H. Peng, T. Barker, C. Q. Xia, and M. J. Clare-Salzler, "Peptide-pulsed immature dendritic cells reduce response to β cell target antigens and protect NOD recipients from type I diabetes," *Annals of the New York Academy of Sciences*, vol. 1079, no. 1, pp. 153–156, 2006.

[14] M. A. Ruffner and P. D. Robbins, "Dendritic cells transduced to express interleukin 4 reduce diabetes onset in both normoglycemic and prediabetic nonobese diabetic mice," *PLoS One*, vol. 5, no. 7, article e11848, 2010.

[15] R. J. Creusot, P. Chang, D. G. Healey, I. Y. Tcherepanova, C. A. Nicolette, and C. G. Fathman, "A short pulse of IL-4 delivered by DCs electroporated with modified mRNA can both prevent and treat autoimmune diabetes in NOD mice," *Molecular Therapy*, vol. 18, no. 12, pp. 2112–2120, 2010.

[16] E. van Etten, O. Dardenne, C. Gysemans, L. Overbergh, and C. Mathieu, "1,25-Dihydroxyvitamin D_3 alters the profile of bone marrow-derived dendritic cells of NOD mice," *Annals of the New York Academy of Sciences*, vol. 1037, no. 1, pp. 186–192, 2004.

[17] J. Machen, J. Harnaha, R. Lakomy, A. Styche, M. Trucco, and N. Giannoukakis, "Antisense oligonucleotides down-regulating costimulation confer diabetes-preventive properties to nonobese diabetic mouse dendritic cells," *The Journal of Immunology*, vol. 173, no. 7, pp. 4331–4341, 2004.

[18] J. Morin, B. Faideau, M. C. Gagnerault, F. Lepault, C. Boitard, and S. Boudaly, "Passive transfer of flt-3L-derived dendritic cells delays diabetes development in NOD mice and associates with early production of interleukin (IL)-4 and IL-10 in the spleen of recipient mice," *Clinical and Experimental Immunology*, vol. 134, no. 3, pp. 388–395, 2003.

[19] J. Tian, A. P. Olcott, and D. L. Kaufman, "Antigen-based immunotherapy drives the precocious development of autoimmunity," *The Journal of Immunology*, vol. 169, no. 11, pp. 6564–6569, 2002.

[20] J. Tian, A. P. Olcott, L. R. Hanssen, D. Zekzer, B. Middleton, and D. L. Kaufman, "Infectious Th1 and Th2 autoimmunity in diabetes-prone mice," *Immunological Reviews*, vol. 164, no. 1, pp. 119–127, 1998.

[21] P. V. Lehmann, E. E. Sercarz, T. Forsthuber, C. M. Dayan, and G. Gammon, "Determinant spreading and the dynamics of the autoimmune T-cell repertoire," *Immunology Today*, vol. 14, no. 5, pp. 203–208, 1993.

[22] R. Tisch, X. D. Yang, S. M. Singer, R. S. Liblau, L. Fugger, and H. O. McDevitt, "Immune response to glutamic acid decarboxylase correlates with insulitis in non-obese diabetic mice," *Nature*, vol. 366, no. 6450, pp. 72–75, 1993.

[23] J. Tian, S. Gregori, L. Adorini, and D. L. Kaufman, "The frequency of high avidity T cells determines the hierarchy of determinant spreading," *The Journal of Immunology*, vol. 166, no. 12, pp. 7144–7150, 2001.

[24] D. L. Kaufman, M. Clare-Salzler, J. Tian et al., "Spontaneous loss of T-cell tolerance to glutamic acid decarboxylase in murine insulin-dependent diabetes," *Nature*, vol. 366, no. 6450, pp. 69–72, 1993.

[25] J. Tian and D. L. Kaufman, "Attenuation of inducible Th2 immunity with autoimmune disease progression," *The Journal of Immunology*, vol. 161, no. 10, pp. 5399–5403, 1998.

[26] A. P. Olcott, J. Tian, V. Walker et al., "Antigen-based therapies using ignored determinants of β cell antigens can more effectively inhibit late-stage autoimmune disease in diabetes-prone mice," *The Journal of Immunology*, vol. 175, no. 3, pp. 1991–1999, 2005.

[27] A. Cadili and C. de Gara, "Complications of splenectomy," *The American Journal of Medicine*, vol. 121, no. 5, pp. 371–375, 2008.

[28] S. N. Mueller and R. Ahmed, "Lymphoid stroma in the initiation and control of immune responses," *Immunological Reviews*, vol. 224, no. 1, pp. 284–294, 2008.

[29] C. King, A. Ilic, K. Koelsch, and N. Sarvetnick, "Homeostatic expansion of T cells during immune insufficiency generates autoimmunity," *Cell*, vol. 117, no. 2, pp. 265–277, 2004.

[30] D. V. Baev, S. Caielli, F. Ronchi et al., "Impaired SLAM-SLAM homotypic interaction between invariant NKT cells and dendritic cells affects differentiation of IL-4/IL-10-secreting NKT2 cells in nonobese diabetic mice," *The Journal of Immunology*, vol. 181, no. 2, pp. 869–877, 2008.

[31] K. Takahashi, J. Satoh, and Y. Oka, "Lowered expressions of the NF-κB family members in dendritic cells from NOD mice are associated with a reduced expression of GATA-2," *Annals of the New York Academy of Sciences*, vol. 1150, no. 1, pp. 59-60, 2008.

[32] J. Morin, A. Chimenes, C. Boitard, R. Berthier, and S. Boudaly, "Granulocyte-dendritic cell unbalance in the non-obese diabetic mice," *Cellular Immunology*, vol. 223, no. 1, pp. 13–25, 2003.

[33] R. Peng, K. Bathjat, Y. Li, and M. J. Clare-Salzler, "Defective maturation of myeloid dendritic cell (DC) in NOD mice is controlled by IDD10/17/18," *Annals of the New York Academy of Sciences*, vol. 1005, no. 1, pp. 184–186, 2003.

[34] S. Boudaly, J. Morin, R. Berthier, P. Marche, and C. Boitard, "Altered dendritic cells (DC) might be responsible for regulatory T cell imbalance and autoimmunity in nonobese diabetic (NOD) mice," *European Cytokine Network*, vol. 13, no. 1, pp. 29–37, 2002.

[35] K. Schuster, J. Gadiot, R. Andreesen, A. Mackensen, T. F. Gajewski, and C. Blank, "Homeostatic proliferation of naïve $CD8^+$ T cells depends on CD62L/L-selectin-mediated homing to peripheral LN," *European Journal of Immunology*, vol. 39, no. 11, pp. 2981–2990, 2009.

[36] J. Suffner, K. Hochweller, M. C. Kuhnle et al., "Dendritic cells support homeostatic expansion of $Foxp3^+$ regulatory T cells in Foxp3.LuciDTR mice," *The Journal of Immunology*, vol. 184, no. 4, pp. 1810–1820, 2010.

[37] S. R. Jacobs, R. D. Michalek, and J. C. Rathmell, "IL-7 is essential for homeostatic control of T cell metabolism in vivo," *The Journal of Immunology*, vol. 184, no. 7, pp. 3461–3469, 2010.

[38] O. Boyman, S. Letourneau, C. Krieg, and J. Sprent, "Homeostatic proliferation and survival of naïve and memory T cells," *European Journal of Immunology*, vol. 39, no. 8, pp. 2088–2094, 2009.

[39] A. Ali, M. Garrovillo, M. X. Jin, M. A. Hardy, and S. F. Oluwole, "Major histocompatibility complex class I peptide-pulsed host dendritic cells induce antigen-specific acquired thymic tolerance to islet cells," *Transplantation*, vol. 69, no. 2, pp. 221–226, 2000.

[40] M. Garrovillo, A. Ali, H. A. Depaz et al., "Induction of transplant tolerance with immunodominant allopeptide-pulsed host lymphoid and myeloid dendritic cells," *American Journal of Transplantation*, vol. 1, no. 2, pp. 129–137, 2001.

[41] F. Fu, Y. Li, S. Qian et al., "Costimulatory molecule-deficient dendritic cell progenitors (MHC class II^+, CD80dim, $CD86^-$) prolong cardiac allograft survival in non-immunosuppressed recipients," *Transplantation*, vol. 62, no. 5, pp. 659–665, 1996.

[42] M. B. Lutz, R. M. Suri, M. Niimi et al., "Immature dendritic cells generated with low doses of GM-CSF in the absence of IL-4 are maturation resistant and prolong allograft survival *in vivo*," *European Journal of Immunology*, vol. 30, no. 7, pp. 1813–1822, 2000.

[43] M. J. Clare-Salzler, J. Brooks, A. Chai, K. Van Herle, and C. Anderson, "Prevention of diabetes in nonobese diabetic mice by dendritic cell transfer," *The Journal of Clinical Investigation*, vol. 90, no. 3, pp. 741–748, 1992.

[44] M. Feili-Hariri, X. Dong, S. M. Alber, S. C. Watkins, R. D. Salter, and P. A. Morel, "Immunotherapy of NOD mice with bone marrow-derived dendritic cells," *Diabetes*, vol. 48, no. 12, pp. 2300–2308, 1999.

[45] M. Feili-Hariri, D. H. Falkner, A. Gambotto et al., "Dendritic cells transduced to express interleukin-4 prevent diabetes in

nonobese diabetic mice with advanced insulitis," *Human Gene Therapy*, vol. 14, no. 1, pp. 13–23, 2003.

[46] J. Tian, M. Clare-Salzler, A. Herschenfeld et al., "Modulating autoimmune responses to GAD inhibits disease progression and prolongs islet graft survival in diabetes–prone mice," *Nature Medicine*, vol. 2, no. 12, pp. 1348–1353, 1996.

[47] A. Muir, A. Peck, M. Clare-Salzler et al., "Insulin immunization of nonobese diabetic mice induces a protective insulitis characterized by diminished intraislet interferon-gamma transcription," *The Journal of Clinical Investigation*, vol. 95, no. 2, pp. 628–634, 1995.

[48] D. Elias and I. R. Cohen, "Peptide therapy for diabetes in NOD mice," *The Lancet*, vol. 343, no. 8899, pp. 704–706, 1994.

[49] J. F. Elliott, H. Y. Qin, S. Bhatti et al., "Immunization with the larger isoform of mouse glutamic acid decarboxylase (GAD^{67}) prevents autoimmune diabetes in NOD mice," *Diabetes*, vol. 43, no. 12, pp. 1494–1499, 1994.

[50] C. Engman, Y. Wen, W. S. Meng, R. Bottino, M. Trucco, and N. Giannoukakis, "Generation of antigen-specific Foxp3 + regulatory T-cells *in vivo* following administration of diabetes-reversing tolerogenic microspheres does not require provision of antigen in the formulation," *Clinical Immunology*, vol. 160, no. 1, pp. 103–123, 2015.

[51] B. Ahrens, C. Gruber, R. D. Rha et al., "BCG priming of dendritic cells enhances T regulatory and Th1 function and suppresses allergen-induced Th2 function in vitro and in vivo," *International Archives of Allergy and Immunology*, vol. 150, no. 3, pp. 210–220, 2009.

[52] A. W. Thomson, H. R. Turnquist, A. F. Zahorchak, and G. Raimondi, "Tolerogenic dendritic cell-regulatory T-cell interaction and the promotion of transplant tolerance," *Transplantation*, vol. 87, Supplement 9, pp. S86–S90, 2009.

[53] M. Feili-Hariri, D. H. Falkner, and P. A. Morel, "Regulatory Th2 response induced following adoptive transfer of dendritic cells in prediabetic NOD mice," *European Journal of Immunology*, vol. 32, no. 7, pp. 2021–2030, 2002.

[54] K. V. Tarbell, S. Yamazaki, K. Olson, P. Toy, and R. M. Steinman, "$CD25^+$ $CD4^+$ T cells, expanded with dendritic cells presenting a single autoantigenic peptide, suppress autoimmune diabetes," *The Journal of Experimental Medicine*, vol. 199, no. 11, pp. 1467–1477, 2004.

[55] S. Yamazaki, K. Inaba, K. V. Tarbell, and R. M. Steinman, "Dendritic cells expand antigen-specific $Foxp3^+CD25^+CD4^+$ regulatory T cells including suppressors of alloreactivity," *Immunological Reviews*, vol. 212, no. 1, pp. 314–329, 2006.

[56] S. Yamazaki, T. Iyoda, K. Tarbell et al., "Direct expansion of functional $CD25^+$ $CD4^+$ regulatory T cells by antigen-processing dendritic cells," *The Journal of Experimental Medicine*, vol. 198, no. 2, pp. 235–247, 2003.

[57] S. Yamazaki, M. Patel, A. Harper et al., "Effective expansion of alloantigen-specific $Foxp3^+$ $CD25^+$ $CD4^+$ regulatory T cells by dendritic cells during the mixed leukocyte reaction," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 8, pp. 2758–2763, 2006.

[58] S. Gregori, M. Battaglia, and M. G. Roncarolo, "Re-establishing immune tolerance in type 1 diabetes via regulatory T cells," *Novartis Foundation Symposium*, vol. 292, pp. 174–183, 2008, discussion 183-176, 202-173.

[59] K. Mahnke, E. Schmitt, L. Bonifaz, A. H. Enk, and H. Jonuleit, "Immature, but not inactive: the tolerogenic function of immature dendritic cells," *Immunology and Cell Biology*, vol. 80, no. 5, pp. 477–483, 2002.

[60] S. Adams, D. W. O'Neill, and N. Bhardwaj, "Recent advances in dendritic cell biology," *Journal of Clinical Immunology*, vol. 25, no. 2, pp. 87–98, 2005.

[61] H. Jonuleit, E. Schmitt, K. Steinbrink, and A. H. Enk, "Dendritic cells as a tool to induce anergic and regulatory T cells," *Trends in Immunology*, vol. 22, no. 7, pp. 394–400, 2001.

[62] H. Jonuleit, E. Schmitt, G. Schuler, J. Knop, and A. H. Enk, "Induction of interleukin 10–producing, nonproliferating $CD4^+$ T cells with regulatory properties by repetitive stimulation with allogeneic immature human dendritic cells," *The Journal of Experimental Medicine*, vol. 192, no. 9, pp. 1213–1222, 2000.

[63] K. Steinbrink, M. Wolfl, H. Jonuleit, J. Knop, and A. H. Enk, "Induction of tolerance by IL-10-treated dendritic cells," *The Journal of Immunology*, vol. 159, no. 10, pp. 4772–4780, 1997.

[64] V. Kronin, L. Wu, S. Gong, M. C. Nussenzweig, and K. Shortman, "DEC-205 as a marker of dendritic cells with regulatory effects on CD8 T cell responses," *International Immunology*, vol. 12, no. 5, pp. 731–735, 2000.

[65] A. M. D'Alise, V. Auyeung, M. Feuerer et al., "The defect in T-cell regulation in NOD mice is an effect on the T-cell effectors," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 50, pp. 19857–19862, 2008.

[66] J. Lo, *Dendritic Cell Therapy for Type 1 Diabetes: Role of Antigen-Presentation in Immune Modulation (Doctoral Disseration)*, University of Florida, Gainesville, FL, USA, 2009.

Hindawi
Journal of Immunology Research
Volume 2018, Article ID 7685371, 13 pages
https://doi.org/10.1155/2018/7685371

Research Article

Frequency of Interferon-Resistance Conferring Substitutions in Amino Acid Positions 70 and 91 of Core Protein of the Russian HCV 1b Isolates Analyzed in the T-Cell Epitopic Context

V. S. Kichatova,[1,2,3] **K. K. Kyuregyan**,[1,2,3] **N. V. Soboleva**,[1] **A. A. Karlsen**,[1,2,3,4] **O. V. Isaeva**,[2,3] **M. G. Isaguliants**,[1,4,5] **and M. I. Mikhailov**[2,3]

[1]*Chumakov Federal Scientific Center for Research and Development of Immune-and-Biological Products of Russian Academy of Sciences, Moscow 108819, Russia*
[2]*Russian Medical Academy of Continuous Professional Education, Moscow 125993, Russia*
[3]*Mechnikov Research Institute for Vaccines and Sera, Moscow 105064, Russia*
[4]*NF Gamaleja Research Center of Epidemiology and Microbiology, Moscow 123098, Russia*
[5]*Riga Stradins University, Riga LV-1007, Latvia*

Correspondence should be addressed to V. S. Kichatova; vera_kichatova@mail.ru
and K. K. Kyuregyan; karen-kyuregyan@yandex.ru

Received 29 August 2017; Accepted 29 November 2017; Published 7 February 2018

Academic Editor: Pedro A. Reche

Amino acid substitutions R70Q/H and L91M in HCV subtype 1b core protein can affect the response to interferon and are associated with the development of hepatocellular carcinoma. We found that the rate of R70Q/H in HCV 1b from Russia was 31.2%, similar to that in HCV strains from Asia (34.0%), higher than that in the European (18.0%, $p = 0.0010$), but lower than that in the US HCV 1b strains (62.8%, $p < 0.0001$). Substitution L91M was found in 80.4% of the Russian HCV 1b isolates, higher than in Asian isolates (43.8%, $p < 0.0001$). Thus, a significant proportion of Russian HCV 1b isolates carry the unfavorable R70Q/H and/or L91M substitution. In silico analysis of the epitopic structure of the regions of substitutions revealed that both harbor clusters of T-cell epitopes. Peptides encompassing these regions were predicted to bind to a panel of HLA class I molecules, with substitutions impairing peptide recognition by HLA I molecules of the alleles prevalent in Russia. This indicates that HCV 1b with R70Q/H and L91M substitutions may have evolved as the immune escape variants. Impairment of T-cell recognition may play a part in the negative effect of these substitutions on the response to IFN treatment.

1. Introduction

Approximately 177.5 million people worldwide are infected with the hepatitis C virus (HCV). In 60–80% of cases, infection results in the chronic liver disease; 10 to 25% progress to cirrhosis and hepatocellular carcinoma (HCC) [1–4], the second most common cause of cancer death worldwide after the lung cancer [5]. HCV infection is associated with at least half of the HCC cases [6]. In patients with hepatitis C, treatment with the direct-acting antivirals (DAAs) of the second generation results in the sustained virological response (SVR) in 98-99% cases. Unfortunately, due to the high cost and limited availability of DAAs, clinical practice in low- to moderate-income countries still relies on the pegylated interferon and ribavirin (PEG-IFN/RBV) therapy.

Infections with HCV genotype 1 that still predominate (in 48% of HCV-infected people), specifically HCV 1b

(in 23% of HCV-infected people [4]), have been repeatedly associated with poor prognosis [7–11]. In chronic hepatitis C, infection with HCV 1b increases the rate of progression to fibrosis and cirrhosis and the risk of HCC development compared to infections with other HCV genotypes [8–12]. HCV genotype 1 infection is also an important prognostic factor of poor SVR to PEG-IFN/RBV treatment, with the response rate for genotypes 2 and 3 reaching 80%, and for genotype 1, only 35 to 45% [13–15], being lower or tending to be lower for HCV 1b carriers (37% of patients with subtype 1b and 45% of those with subtype 1a [16]). Several studies and meta-analyses have concluded that eradication of HCV with antiviral therapy reduces the risk of HCC in patients with chronic hepatitis C, but the risk is not eliminated [17]. Wide spread of HCV 1b, high rate of progression to chronic hepatitis C, poor response to PEG-IFN/RBV therapy still actual in poor to modest healthcare settings, and the increased risks of HCC development prompt continuation of the studies on the mechanisms of HCV 1b resistance to interferon treatment.

Akuta et al. were the first to report that amino acid substitutions in positions 70 and 91 of the HCV 1b core protein associate with virological response to PEG-IFN/RBV treatment. They found that change of arginine to glutamine or histidine in amino acid (aa) position 70 (R70Q and R70H or R70Q/H) and/or methionine in aa position 91 (L91M) was present in 100% of nonresponders who tested positive for HCV RNA at the end of 48 weeks of PEG-IFN/RBV treatment, but in only 42% of responders [18]. The significance of R70Q for HCV 1b (and also for subtype 5a) was confirmed in a series of later studies (see [19] for a review). R70Q was recorded as a predictive marker for the resistance to PEG-IFN/RBV also in the prolonged treatment regimens (72 weeks) and in the triple therapy with PEG-IFN/RBV and second-generation DAA [19–21]. A series of studies determined the frequency of these mutations in different population groups [22–24]. All, including the very recent ones, concluded that the nature of amino acid residues in aa position 70 of HCV core can help to distinguish patients who can still benefit from the affordable IFN-based therapy from those who must be treated with DAAs to prevent the evolution towards the end-stage liver disease [25]. The clinical significance of substitutions in aa position 91 of HCV 1b is less well defined [19] and for other HCV genotypes either unclear [19] or not demonstrated [26–29]. Interestingly, clinical and experimental studies demonstrated the involvement of both of these mutations in hepatocellular carcinogenesis [30, 31]. Amino acid substitution R70Q/H appeared to be associated with cirrhosis and development of HCC even in PEG-IFN/RBV-treated patients achieving SVR [17, 22, 32–38]. Altogether, this motivates the need to characterize polymorphisms in HCV 1b core protein and identify patients carrying "carcinogenic" HCV 1b variants staying at a high risk of developing HCC even after the successful treatment.

The prevalence of HCV core polymorphisms R70Q/H and/or L91M in the territory of the Russian Federation was not characterized despite high incidence of HCV infections [39], the predominance of HCV 1 genotype, specifically of HCV 1b subtype (up to 50% [40–42]), and wide use of standard interferon therapy. In this study, we filled this gap and described the spread in the Russian Federation of HCV 1b strains with R70Q/H and/or L91M substitutions in comparison to their spread in the other geographical regions and evaluated the effect of these substitutions on the recognition of these regions by the immune system of the patients.

2. Materials and Methods

2.1. Samples. The analysis included 313 sequences encoding HCV core amplified from the serum samples of patients with chronic hepatitis C collected during 2007–2014 in different regions of the Russian Federation: Moscow and the Moscow region ($n = 238$), Rostov/Rostov-on-Don ($n = 12$), Tuva/Kyzyl ($n = 12$), Khabarovsk ($n = 36$), Sverdlovsk/Yekaterinburg ($n = 4$), and Sakha/Yakutsk ($n = 11$). Age and gender were known for 117 individuals; 34 were females and 83 were males (the female to male ratio was 1 : 2.4). Mean age was 40 ± 14 years. All serum samples included into the database were from the general population with the exception of those from Moscow/Moscow region represented by the subgroup of the general population without known risk factors ($n = 49$) and intravenous drug users (IDUs; $n = 45$). Written informed consent was obtained from each patient.

2.2. HCV Genotyping and Sequencing. HCV genotype and subtype were determined by the phylogenetic analysis of HCV core coding region; and for HCV 2k/1b recombinants, also of the NS5b region. Nucleic acid extraction was performed using QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany), or MagNA Pure Compact Nucleic Acid Isolation Kit I (Roche Applied Science, Mannheim, Germany), or Sileks MagNA kit (Sileks, Moscow, Russia). Reverse transcription and amplification was performed with Transcriptor First Strand cDNA Synthesis Kit and Fast Start High Fidelity PCR System (Roche Applied Science). HCV core coding region (nucleotide positions 273–1315 according to HCV 1a reference strain H77, GenBank accession number AF011753) was subjected to nested PCR with the following primers: outer forward 5′-GCT AGC CGA GTA GTG TTG GG-3′; outer reverse 5′-ACC AGT TCA TCATCA TAT YCC-3′; inner forward 5′-GAA AGG CCT TGT GGT ACT GC-3′; and inner reverse 5′-TTC ATC ATA TTC CAT GCCA-3′. The first and the second steps of PCR were as follows: 5 min at 94°C, then 35 cycles of 45 sec at 94°C, 45 sec at 55°C, 90 sec at 72°C, and final elongation for 7 min at 72°C. PCR products of 1043 nt length were cut and extracted from agarose gel using QIAquick Gel Extraction kit (QIAGEN). Sequencing was performed on an automated sequencer 3130 Genetic Analyzer (ABI, Foster City, USA) using Big Dye Terminator v.3.1 Cycle Sequencing Kit according to the manufacturer's protocol. The genotype was established by comparison sequences

with GenBank reference sequences representative for all HCV genotypes.

2.3. HCV Sequence Database. Database of 313 HCV sequences was supplemented with 94 sequences from GenBank geographically attributed to St. Petersburg ($n = 42$: AY070167–AY070169, AY070174.1, AY070178–AY070211, AY070214, AY070215, AY587844, and AY587845); Novosibirsk ($n = 47$: DQ001223–DQ001264, DQ001267–DQ001270, AH014196.2, and AH014197.2); and Moscow ($n = 5$: AF176573, KM054515, KM054516, KT983617, and X71407). The complete database consisted of 407 sequences of the "Russian" HCV core of subtypes: 1a ($n = 29$), 1b ($n = 189$), 3a ($n = 171$), and recombinant 2k/1b ($n = 18$) (Table 1). Group of comparison consisted of 2210 HCV core sequences from the Los Alamos database (LANL; https://hcv.lanl.gov/content/index) distributed as follows: Asia ($n = 374$), USA ($n = 1343$), and Europe ($n = 497$) (Table 1). Samples of the Russian patients infected with HCV1b formed a separate set divided into the subsets depending on the region of sample collection (Table 2). The geographical distribution of these subsets is illustrated by Supplementary Figure S1.

2.4. In Silico Analysis of HCV Core Sequences. Alignment of the nucleotide and predicted amino acid sequences of HCV were performed using MEGA 7.0.18. Prevalence of substitutions in amino acid positions 70 and 91 was calculated using Microsoft Office Excel. Phylogenetic analysis of HCV1b core sequences (Supplementary Figure S2) was done using a Bayesian likelihood-based algorithm implemented in Beast version 1.8.4 [43]. The SRD06 substitution model was used with a relaxed lognormal clock. Analysis was run over 700 million generations and trees were sampled every 70,000 generations, resulting in 70,000 trees. Trees were annotated with TreeAnnotator v.1.8.3 using a burn-in of 1000 trees and visualized with FigTree v.1.4.2 (Andrew Rambaut; http://tree.bio.ed.ac.uk/software/figtree/). Search for T-cell epitopes in HCV core sequences was performed through EIDB database (http://www.iedb.org) [44]. Epitope sequences were aligned with HCV core aa sequences using MEGA software.

2.5. Prediction of Presentation of HCV Core-Derived Peptides by HLA Class I and II Molecules. Amino acid sequences of HCV core of 189 HCV 1b were used to build the consensus sequence of "Russian" HCV 1b core (sequence in GenBank deposition). Consensus sequence was divided into nine amino acid- (aa-) long peptides with mandatory presence of aa residues 70 or 91. Prediction of presentation of resulting HCV core-derived peptides by human leukocyte antigen class I molecules (HLA class I) was done on EPITOPE VACCINE OPTIMIZATION server (http://bio.med.ucm.es/episopt.html) using EPISPOT tool (http://bio.med.ucm.es/episopt.html; [45]). Analysis was carried for a selection of HLA I alleles prevalent in the Russian population based on the most representative dataset [46] and other data deposited in the Allele Frequency Net Database (http://www.allelefrequencies.net/pop6001c.asp?pop_id=3322; [47]).

Presentation by HLA class II molecules was evaluated using the NetMHCIIpan 3.1 program (http://www.cbs.dtu.dk/services/NetMHCIIpan-3.1) [48–50] for the region starting with 20-mer peptide with the C-terminal aa 70 and ending with the 20-mer peptide with the N-terminal aa 91. Analysis was done for the following: 275 HLA class II alleles DRB1*1101–DRB1*1195, DRB1*1301–DRB1*1399, and DRB1*1501–DRB1*1549; and combination of alpha and beta chains DQA1*0101, DQA1*0102, DQA1*0301, DQA1*0501 with DQB1*0201, DQB1*0301, DQB1*0501, DQB1*0602, DQB1*0603, DQB1*0604, DQB1*0605, DQB1*0606, DQB1*0607, and DQB1*0608, selected based on the published data on the frequency of HLA class II alleles in the Russian Federation [46, 47, 51]. Strong binding was characterized by binding level < 2%; weak binding, 2 to 10%; no binding, >10%.

2.6. Statistical Analysis. Data analysis was performed using the http://Graphpad.com/. Statistical significance was evaluated by Fisher's exact test using parametric model, two-tailed p value < 0.05 was considered statistically significant.

3. Results

3.1. Characteristics of the Set of HCV 1b Sequences Used in the Study. The analysis included 313 sequences encoding HCV core amplified from the serum samples of patients with chronic hepatitis C collected during 2007–2014 from the general population in different regions of the Russian Federation, supplemented with 94 sequences from the GenBank ($n = 407$; Table 1). Sequence set included 189 HCV 1b sequences derived from the territory of the Russian Federation, 140 obtained here, and 49 sequences from the GenBank. We performed their phylodynamic analysis with the focus on 140 HCV 1b isolates (Supplementary Figure S2). The analysis illustrated a long history of HCV 1b evolution on the Russian territory, with current strains having an average time of separation from the "foreign" strains of approximately 15 ± 5 years, the oldest strains separating 60 ± 15 years ago (Supplementary Figure S2). No clusters were revealed, indicating that sequences were not related. HCV 1b sequences with R70Q/H and/or L91M substitutions did not form any clusters either, being distributed evenly throughout the phylogenetic tree (Supplementary Figure S2). This demonstrates that the selected sequences are geographically genuine, adequately reproduce the diversity of HCV 1b in the territory of the Russian Federation, and are, therefore, suitable for the analysis of frequency of occurrence of amino acid substitutions in positions 70 and 91 in the territory of the Russian Federation.

3.2. Frequency of Substitutions in aa Positions 70 and 91 of HCV 1b from the Russian Federation Compared to Other Geographical Regions. Substitution R70Q was found in 27.0% and R70H in 4.2% of the Russian HCV 1b isolates, which was similar to their frequencies in the isolates from Asia and in all countries taken together ($p > 0.05$), but significantly higher than the frequency of R70Q and of R70H in

Table 1: Frequency of occurrence of substitutions at amino acid positions 70 and 91 of the nucleocapsid (core) protein of HCV genotypes 1a, 1b, and 3a in different population groups (in %).

HCV genotype	Substitutions	Geographical variation in % of sequences variants in HCV of different genotypes				
		Russia	Europe	Asia	USA	Total
1a	nn samples	$n = 29$	$n = 149$	$n = 27$	$n = 1087$	$n = 1292$
	Q70	3.4	0.7	0	1.7	1.5
	H70	0	0	0	0.1	0.1
	M91	0	0	0	0	0
	Both*	0	0	0	0	0
	None**	96.6	99.3	100	98.2	98.4
1b	nn samples	$n = 189$	$n = 316$	$n = 224$	$n = 218$	$n = 947$
	Q70	27.0	17.7	31.3	59.6	32.4
	H70	4.2	0.3	2.7	3.2	2.3
	M91	80.4	85.8	43.8	86.2	74.9
	Both*	27.5	15.5	16.1	54.6	27.0
	None**	15.9	11.7	38.4	5.5	17.4
3a	nn samples	$n = 171$	$n = 28$	$n = 123$	$n = 38$	$n = 360$
	Q70	2.9	7.1	9.8	0	5.3
	H70	0	0	0	0	0
	M91	0.6	0	0	0	0.3
	Both*	0	0	0	0	0
	None**	96.5	92.9	90.2	100	94.4

*Substitutions in both aa 70 and 91; **no substitutions.

Europe (17.7%, $p = 0.0179$; and 0.3%, $p = 0.0021$, resp.; Table 1). Substitution L91M was found in 80.4% of HCV 1b isolates from Russia, likewise to Europe and worldwide, but more often than in the strains from Asia (43.8%, $p < 0.0001$; Table 1). Substitution R70Q was significantly more prevalent compared to substitution R70H in all geographical regions including Russia. Frequency of occurrence of substitutions in positions 70 and 91 in HCV genotypes 1a and 3a was low (Table 1). Geographical distribution of amino acid variants in positions 70 and 91 for HCV subtypes 1a and 3a was similar, except for comparatively more frequent occurrence of R70Q in HCV 3a strains from Asia (significantly more than in the Russian HCV subtype 3a strains; 9.8% versus 2.9%, $p = 0.0204$; Table 1). All available core sequences of the recombinant HCV 2k/1b (18 from Russia and 4 from Europe) had no substitutions in either position 70 or 91 (data not shown).

3.3. Pattern of Substitutions in aa Positions 70 and 91 in HCV 1b Strains in the Russian Federation. HCV 1b variants with substitutions R70Q/H and L91M were evenly distributed in all eight regions of the Russian Federation assessed (Table 2). Differences were observed only between Moscow (R70Q/H, 40.4%; L91M, 79.8%) and Novosibirsk subcohorts (R70Q/H, 20.7%, $p = 0.0049$; L91M, 75.5%, $p = 0.0094$). Substitution R70Q dominated over substitution R70H in all studied cohorts (data not shown). Prevalence of R70Q/H and L91M mutations were similar in males and females ($p = 0.8286$ and $p = 0.439$ for aa positions 70 and 91, resp.). Comparison of subcohorts with a known year of sample collection demonstrated the accumulation of R70Q/H substitution over time: they were found in 19.6% of samples collected before 2005 and in 39.1% in samples collected in 2011–2014 ($p = 0.038$; Table 2). An increase in the frequency of occurrence of the substitution was observed only for position 70 but not for 91 (Figure 1).

Analysis of the Moscow cohort demonstrated similar frequency of R70Q/H and L91M mutations among the general population and the intravenous drug users (IDUs) (R70Q/H: 38.7% versus 42.2%, $p = 0.8341$; L91M: 81.6% versus 77.7%, $p = 0.7979$; Table 2). Difference between these two population groups was observed only in aa residue in position 75: R75H/Q occurred in the general population more frequently than in IDUs (52.6% and 46.6%, resp.; $p = 0.0423$). Substitutions at aa position 75 were repeatedly detected in the earlier studies [52, 53], but their clinical significance is yet unclear.

3.4. Epitopic Analysis of the Region of HCV Core Containing Amino Acid Residues 70 and 91. Amino acid sequences of HCV core of 189 HCV 1b were used to build the consensus sequence of the "Russian" HCV 1b core, an "average" HCV sequence largely free from the patient-specific adaptations [54]. Epitopic analysis of the consensus HCV core using Immune Epitope Database and Analysis resource (http://www.iedb.org) identified 183 CD4+ and CD8+ T cell epitopes, with 14 (7.7%) encompassing aa 70 and 29 (15.8%) encompassing aa 91, respectively. Totally, 23.5%

Table 2: Occurrence of substitutions in amino acid positions 70 and 91 of the nucleocapsid (core) protein of HCV 1b in the samples collected from the general population of the Russian Federation during the routine blood tests.

Number of cohort	Federal District of Russia	City	Date of collection of samples	Number of HCV 1b sequences	Mutations found in aa position 70, *n* (%)	Mutations found in aa position 91, *n* (%)	Mutations found in aa positions 70 and 91, *n* (%)	Without mutations in aa positions 70 and 91, *n* (%)
1	Central	Moscow	2008	5	1 (20.0)	4 (80.0)	1 (20.0)	1 (20.0)
			2009	6	1 (16.7)	5 (83.3)	0 (0)	0 (0)
			2011	17	10 (58.9)	15 (88.2)	9 (53.0)	1 (5.9)
			2014	21	7 (33.3)	16 (76.2)	6 (28.6)	4 (19.0)
			2008–2014	49	19 (38.7)	40 (81.6)	17 (34.7)	7 (14.2)
			2014	45*	19 (42.2)	35 (77.7)	16 (35.6)	7 (15.6)
			2008–2014	94**	38 (40.4)	75 (79.8)	33 (35.1)	14 (14.9)
2	Northwest	Saint Petersburg	Before 2001	18	4 (22.2)	15 (83.3)	3 (16.7)	2 (11.1)
3	Southern	Rostov-on-Don	2008	8	0 (0)	8 (100)	0 (0)	0 (0)
4	Ural	Yekaterinburg	2008	3	1 (33.3)	3 (100)	1 (33.3)	0 (0)
5	Siberian	Novosibirsk	Before 2005	29	6 (20.8)	22 (75.9)	6 (20.7)	7 (23.8)
6		Kyzyl	2008	6	1 (16.7)	4 (66.7)	0 (0)	1 (16.7)
7	Far Eastern	Yakutsk	2008	7	1 (14.3)	6 (85.7)	1 (14.3)	1 (14.3)
8		Khabarovsk	2009	24	8 (33.3)	19 (79.2)	8 (33.3)	5 (20.8)

*Intravenous drug users (IDUs); **all samples including IDUs.

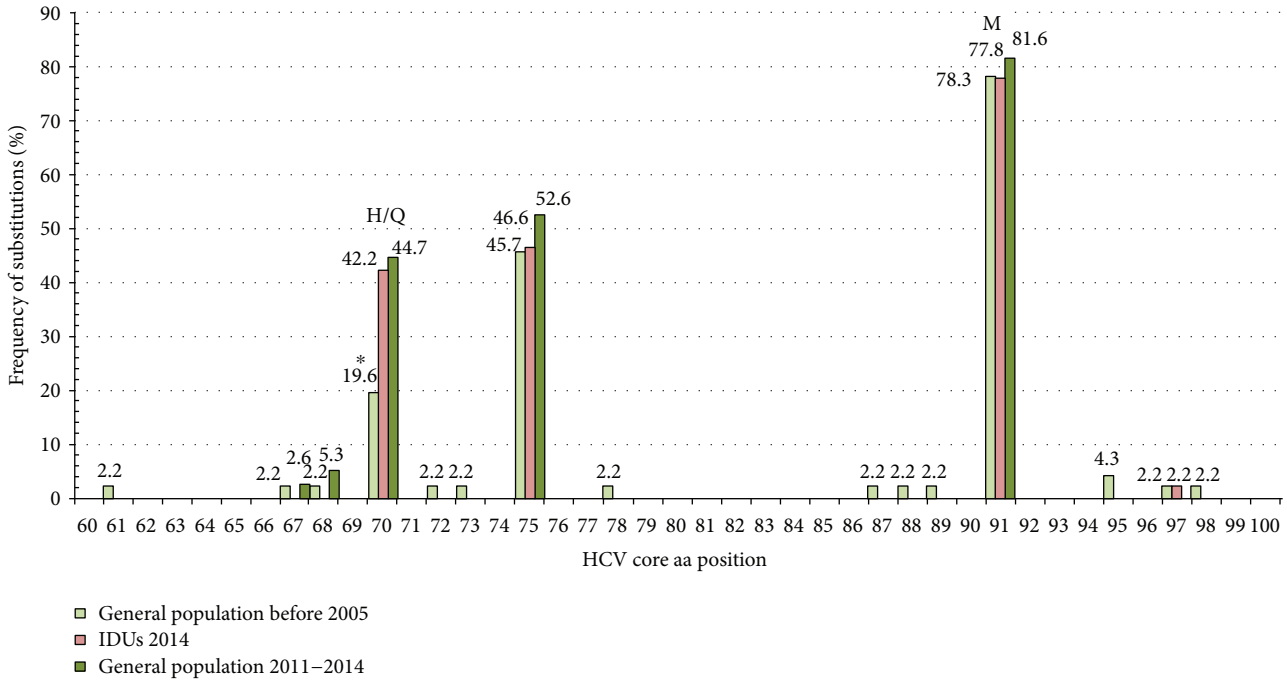


Figure 1: Frequency of amino acid substitutions in the region between aa 60 to 100 of the nucleocapsid (core) of HCV 1b, in samples collected before 2005 in the general population in Russia ("general population before 2005"; $n = 46$), during 2011–2014 in Moscow/Moscow region in the general population ("general population 2011–2014"; $n = 38$), and in intravenous drug users ("IDUs 2014"; $n = 45$). *Frequency of occurrence of R70Q/H in the group "general population before 2005" (19.6%) was significantly lower compared to "general population 2011–2014" (44.7%, $p = 0.0178$) and "IDUs 2014" (42.2%, $p = 0.0240$), while frequency of the substitution in the latter two groups did not differ.

of all T-cell epitopes within HCV core included either aa position 70 or 91. HCV 1b core was found to harbor 8 epitopes encompassing aa 70 and 11 encompassing aa 91 (Supplementary Figure S3).

3.5. Prediction of Recognition of the Region of HCV 1b Core Containing Amino Acid Residues 70 and 91 by HLA Class I and II Molecules. First, we analyzed how HCV core variants with and without mutations were recognized by HLA class I molecules. The complete list of HLA I alleles and peptides within HCV 1b core capable to bind to them is presented in Supplementary Table S1 for regions aa 62 to 78 and S2 for aa 83 to 99. From these lists, we selected HLA I alleles prevalent in the territory of the Russian Federation [28, 29] (Figure 2). Several alleles predicted to bind peptides with the wild type, but not with the mutant amino acid residues, were identified. For peptides encompassing aa position 70, these were A3101, B0702, and B1516 and for aa position 91, A0202, A0205, and B0702 (Figure 2, Tables S1 and S2). HLA-B0702 allele, prevalent in the Russian population (20.7%), turned to be of particular interest as it was predicted to bind peptides containing R70 and L91 but not their mutated variants containing Q70, or H70, and M91 (Figure 2, Tables S1 and S2).

We have also assessed if HCV core variants with and without substitutions can bind to HLA class II molecules. Since HLA class II are capable of binding up to 20 amino acid long peptides, we included in the analysis the consensus amino acid sequence of HCV 1b spanning 39 aa with aa positions 70 or 91 in the middle. Data on the frequency of HLA II alleles in the Russian population with high-resolution typing are limited. Based on the low-resolution analysis for HLA II DR, the most prevalent alleles are HLA II DRB101 (22.4%), DRB104 (20.5%), DRB107 (25.6%), DRB111 (26.7%), DRB113 (26.0%), and DRB115 (23.4%) [28, 29]. High-resolution data on the frequency of HLA II DQA1 and DQB1 alleles for ethnical Russians is available for only one geographical region of the Russian Federation (Astrakhan region) [30]. According to this data, the most prevalent alleles are as follows: DQA1*0101 (24.0%), DQA1*0102 (34.3%), DQA1*0301 (20.0%), DQA1*0501 (47.3%), DQB1*0201 (33.7%), DQB1*0301 (39.3%), DQB1*0501 (20.3%), and DQB1*0602-08 (36.0%) [30]. Analysis of binding of these HLA II molecules with HCV core-derived peptides done using NetMHCIIpan program included the total of 275 alleles. None of the HLA II DRB1 alleles prevalent in Russia were predicted to bind any aa fragment within the consensus HCV 1b core sequence spanning aa positions 50 to 110. Weak binding with the fragment aa 71 to 110 harboring aa 91 was predicted for alleles HLA-DQA1*0501-DQB1*0301, HLA-DQA1*0501-DQB1*0501, HLA-DQA1*0501-DQB1*0603, HLA-DQA1*0501-DQB1*0604, HLA-DQA1*0501-DQB1*0607, HLA-DQA1*0501-DQB1*0608, HLA-DQA1*0301-DQB1*0301, and HLA-DQA1*0301-DQB1*0501 (exemplified by the data set in the Supplementary Table S3). The

Predicted HLA I allele	Epitopes within HCV core encompassing regions of subsitutes					Prevalence based on low resolution HLA I typing	
	aa position 70			aa position 91			
	R70	Q70	H70	L91	M91	HLA I type	Prevalence, %
A0201				*	*		
A0202				*		A02	50,0
A0205				*			
A0209				*	*		
A2402				*	*	A24	20,3
A2902				*	*	A29	3,2
A3101	*					A31	3,9
B0702	*			*		B07	20,7
B0801	*	*	*			B08	12,6
B1508		*					
B1513	*	*	*			B15	11,4
B1516	*						
B2801		*					
B2702	*	*	*				
B2703		*				B27	8,8
B2704		*					
B2705	*						
B2706	*	*	*				
B3501				*	*	B35	20,6
B3909	*	*	*	*	*	B39	4,6
B4402	*	*				B44	18,0
B5101				*	*		
B5102			*	*	*	B51	9,6
B5103					*		
B5301				*	*	B53	0,3
B5401		*				B54	No data
B5502		*	*			B55	1,9
B5701	*	*	*	*	*	B57	5,1
B2702	*	*	*				

Figure 2: Prediction of binding of peptides encompassing aa 70 and 91 of nucleocapsid (core) protein of HCV 1b to HLA I alleles prevalent in the Russian population. Binding prediction was done using EPISPOT tool (http://bio.med.ucm.es/episopt.html [45]). Allele frequency data is based on the data obtained from Allele Frequency Net Database [46, 47]. Predicted binding is depicted by an asterisk. HLA class I molecules binding only to the peptides containing wild-type amino acid residues R70 and/or L91 are shaded green; only variants with Q70, or H70, and/or M91 mutations, red; binding regardless of the nature of amino acid residues in positions 70 and/or 91, blue.

binding was not affected by the nature of amino acid residue in position 91. No binding was predicted for the peptides spanning aa 50 to 90 harboring aa 70 (data not shown).

4. Discussion

The presence of substitutions R to Q/H in aa position 70 of the nucleocapsid (core) of HCV 1b can help distinguish patients who can still benefit from the affordable IFN-based therapy from those who do not respond and are at an increased risk of HCC development [17, 22, 25, 32–38]. There are also indications of the clinical significance of L to M substitution in aa position 91 of HCV 1b, although less well defined [18, 19]. Here, we for the first time analyzed the frequency of occurrence of R70Q/H and L91M substitutions in a set of 189 sequences of HCV 1b isolates derived from the territory of the Russian Federation. Phylogenetic analysis has shown that these sequences formed no clusters, and it evolved as a result of continuous viral circulation in the territory of the former Soviet Union, which indicates that they adequately represent the diversity of Russian HCV 1b isolates. We found that 31.4% of the Russian HCV 1b isolates harbor R70Q/H, that is, almost one-third of individuals infected with HCV 1b in the Russian Federation carry viral isolates that are potentially resistant to the standard PEG-IFN/RBV therapy and confer an increased risk of HCC development. Comparative analysis of HCV 1b core variant distribution demonstrated a higher prevalence of unfavorable R70Q/H variants in the Russian Federation compared to the neighboring European countries. USA was

the only region where the frequency of R70Q mutation appeared to be higher. At the same time, L91M substitution was found to be present in the majority of HCV 1b strains circulating in all regions of the world, including the Russian Federation (except for Asia, where the predominant viral variant had L91). Analysis of the frequency of R70Q/H and L91M in HCV 1b isolated in the different regions of the Russian Federation demonstrated the ubiquitous even distribution of these variants. Example of Moscow/Moscow region has also shown that R70Q/H and L91M HCV 1b variants had similar prevalence in the general population and in the high-risk group of intravenous drug users from the same geographical region, which excluded the influence of this risk group on the results of this study.

We have analyzed the prevalence of substitutions R70Q/H and L91M in all HCV 1b core sequences from LANL ($n = 898$) regardless of the area of sample collection. Similar analysis has been performed by two independent research groups in 2010 [26] and in 2014 [19]. The frequency of occurrence of R70Q and R70H substitutions in HCV 1b isolates in the study done in 2010 was significantly higher than in our analysis (60% and 4% versus 32.9% and 1.7%, resp., for all samples regardless of the time of collection; $p < 0.01$), whereas the frequency of occurrence of L91M substitution was similar (71% in 2010 [26] versus 69% in 2016, $p > 0.05$). Interestingly, we registered an increase in the frequency of occurrence of R70Q/H variant in the Russian Federation between 2005 and 2014: R70Q/H variant was present in 19.6% of samples collected before 2005 and in 39.1% in samples collected in 2011–2014 ($p < 0.05$; Table 2). This pointed at a positive selection of this variant in the Russian population with its prevalence increasing towards the levels observed worldwide.

We observed an increase in the frequency of substitutions in aa position 70, not elsewhere in this region of HCV core, suggesting that this substitution was not related to the changes in the other regions of the core protein. The latter is supported by the absence of covariance of the residues in amino acid position 70 with amino acid residues in other positions of HCV core, or elsewhere in HCV polyprotein [55]. This, together with the long history of evolution of HCV 1b strains in the Russian territory, suggests that substitutions in aa position 70 of HCV 1b core protein are not the compensatory ones. Although, Tasaka-Fujita et al. found polymorphisms at aa 70 to be associated with the efficiency of *in vitro* production of the infectious virus, with deteriorated virus production in 70Q strains resulting in the intracellular accumulation of HCV proteins [30]. Also, a recent study done in 112 Chinese patients with chronic HCV 1b infection revealed that in patients infected with mixtures of 70R and 70Q/H HCV 1b strains (most of the patients in this study), viral kinetics of two strains changed synchronously during the treatment [56]. The latter data indicated that R70Q/H substitutions (i) do not improve viral replication fitness; (ii) have not evolved as a result of the selective pressure of PEG-IFNα/RBV treatment; and (iii) do not confer an advantage in the viral replication/propagation under the treatment conditions [56]. Altogether, these point that positive selection of R70Q/H is unrelated to the viral replication fitness, stressing the importance in the selection of the host-related factors.

Host factors are multiple and include alcohol consumption, age at infection, infections/coinfections, and genetic factors, such as gender, or single-nucleotide polymorphisms upstream of IL-28B [57–60]. Funaoka et al. using a virus culture system demonstrated the suppression of IFN signaling (IFN resistance) in cultured cells infected with 70Q/H HCV 1b strains HCV 1b infected with core mutation (70Q/H), possibly induced by the IL-6-induced upregulation of SOCS3 [61]. HCV 1b-infected patients with 70Q had significantly higher homeostasis model assessment insulin resistance (HOMA-IR) scores compared to HCV 1b patients without substitution in this position, suggesting that the substitution has a close relationship to insulin resistance [62]. Mechanistically, extracellular HCV core protein with substitution at position 70 was found to enhance IL-6 production and reduce adiponectin production from visceral adipose tissue, which can cause insulin resistance, hepatic steatosis, and ultimately development of HCC [63]. This, as well as other experimental and clinical studies identified a series of associations between substitutions in HCV core and such host-related factors as oxidative stress and HCV-induced metabolic disorders [64–66].

An important host genetic factor appeared to be the ethnicity. We have looked for the published data on the links between the response to standard treatment, host ethnicity, and prevalence of core mutations in different population groups. Unusually, poor response to PEG-IFNα/RBV treatment was repeatedly reported in the Afro-American patients compared to Caucasian Americans [67]. People of the African descent had lower chances of success with dual antiviral therapy compared with Caucasians. This was observed in the homogeneous populations with low rates of racial admixture assessed by self-reported ancestry, as well as in admixed populations when ancestry was assessed using genetic markers [68]. American patients of the African origins were found to less likely respond to treatment and achieve SVR than non-African-American patients [69]. Infected with genotype 1, they exhibited significantly lower decreases in the first-phase viral RNA, slower elimination of infected cells, and smaller declines in mean viral RNA over 1 month suggesting an impaired ability to block viral production in African-Americans [70]. On the contrary, Asian patients achieve higher sustained virologic response rates following IFN-based therapy than non-Asians [71]. Interestingly, for the Asian patients, superior virologic outcomes were observed also with different classes of DAAs alone or in combination [71] indicating that the outcomes were not determined by the effects of IFN. These observations fall in line with our findings that Asia is the place where the largest proportion of the virus has not yet acquired both R70Q/H and L91M mutations, whereas in America basically all HCV 1b variants are already mutated (occurrence of nonmutated sequences in America is significantly lower than in Asia, Russia, or Europe; $p < 0.0005$; Table 1).

Also, the treatment outcomes for chronic hepatitis C related to aa substitutions in aa positions 70 and 91 varied

with ethnicity. The first studies demonstrating the importance of R70Q/H and L91M substitutions in the core protein were carried in patients infected with HCV 1b in Japan (18). However, they were not reproduced in all Japanese studies. Enomoto et al. identified the role in treatment response only of the amino acid substitutions in HCV NS5a, but not in the core protein [72]. A study of HCV 1b-infected Caucasian patients in Spain demonstrated a significant association with PEG-IFN/RBV treatment outcome with substitutions R62G, R70Q, and N110T [28]; in Saudi patients, with substitutions R70Q and A75; in Azerbaijani patients from Iran, with substitutions K43R, R70Q, L91M, and S106 [23]. Studies of the Caucasian patients infected with HCV 1b in Sweden, and in Caucasians infected with HCV 1b in the United States demonstrated significant association of PEG-IFN/RBV treatment outcome with substitutions in aa position 70, but not in other aa positions including 91 [26, 27]. Interestingly, despite the unusually poor response to PEG-IFNα/RBV treatment in the Afro-American patients compared to Caucasian Americans, their isolates did not differ in the frequency of occurrence of amino residues in position 70 [27]. Also, the association of differential viral responses with polymorphisms in core aa position 70 demonstrated in the North American patients was weaker than in the Japanese studies [27]. On top of it, in the study of the correlates of PEG-IFNα/RBV treatment response in the Chinese patients with chronic HCV 1b infection, 70Q/H HCV 1b strains exhibited the same virological response as the 70R strains [56]. Analysis of these studies demonstrated that the role of amino acid substitutions in positions other than 70 is discrepant, while substitutions in the position 70 predict poor IFN treatment response in patients of some, but not all ethnicities and nationalities.

While analyzing a broad scope of viral and host factors linking substitutions in aa positions 70 and 91 of HCV core with treatment response, these studies left out the issue of the adaptive immune response, and possible viral evolution under the pressure of the immune system of the host. Immune database analysis done here demonstrated that R70Q/H and L91M substitutions are localized in the T-cell epitopic clusters and may interfere with the immune recognition of host cells infected with mutant virions, which would affect the spontaneous and also IFN therapy-driven immune clearance. Immune recognition in populations of different ethnicity varies; different ethnic groups have distinct and characteristic HLA allele frequencies, resulting in the differential immune recognition of the epitope-carrying regions depending on the haplotype of the host, which would differentially drive viral immune escape. In this context, the increasing prevalence of mutant HCV variants can be explained by an escape from the dominant types of immune pressure in a certain population, in relation to the regionally prevalent HLA types. Such HLA allele-specific mutation patterns were earlier described for HIV-1 [73].

The assumption that HLA allele specific immune pressure would result in the regionally variable patterns of immune escape mutations falls in line with uneven geographical distribution of HCV 1b R70Q/H and L91M variants (Table 1). Furthermore, in silico analysis carried here suggests that at least some of HLA I alleles prevalent in the Russian Federation, such as A02 and B07, are able to bind 70R and 91L peptides but cannot bind peptides with R70Q/H or L91M substitutions. At the same time, the role of HLA class II binding, hence of the T-helper cell or antibody response in immune selection, was predicted as negligible. Due to insufficiency of the high-resolution HLA typing data for the Russian Federation, our predictions of HLA I and HLA II binding with HCV core peptides were done on the low-resolution level. Even with these limitations, our data points at the involvement of HLA I- (but not HLA II-) specific alleles in the differential immune recognition of HCV 1b variants with substitutions in aa positions 70 and 91, with the possibility for immune escape of the variants that have acquired R70Q/H and/or L91M. A very recent comparative analysis of the variation and selection in HCV genome demonstrated positive selection of HCV variants associated with HLA class I-driven CTL response (but not CD4+ T-cell response or RNA structure) [74]. Increasing frequency of occurrence of HCV 1b variants bearing R70Q/H and L91M may result from such positive selection. Interestingly, a study conducted in one particular region of the Russian Federation among patients of the Russian ethnicity found that HLA I haplotypes A02/B07 and A03/B07 are associated with the higher rates of spontaneous HCV elimination and protection against the development of chronic infection [51]. Carriage of A(*)02 was also found to predict SVR in a study done in the Caucasian American patients [75]. In patients with chronic hepatitis C, HLA I allele B07 was related to the resistance to active chronic liver disease indicating its association with recognition and clearance of HCV-infected hepatocytes [51]. These findings support the clinical significance of our observations, requesting further confirmatory studies with high-resolution genotyping of the population.

5. Conclusions

This is the first study characterizing the frequency of occurrence of IFN resistance-conferring mutations in human hepatitis C virus isolates circulating in the territory of the Russian Federation, and the first one, in which the spread of viral variants with substitutions in aa positions 70 and 91 may be associated with the positive selection under the pressure of immune response. Spread of R70Q/H and L91M HCV 1b variants may result from an immune escape from the CTL response. The immunogenetic background of HCV-infected individuals (racial variations in viral-specific immunity) would then determine both the differences in geographical distribution of certain viral variants and ethnical differences in their response to treatment. The choice of treatment strategy for patients with HCV is increasingly based on the personalized approach. It takes into account many viral and host factors. Screening of HCV 1b-infected patients for unfavorable mutations in the core region could be a useful tool to identify the individuals in need of IFN-free treatment regimens employing DAA. In the future, after

complete implementation of IFN-free regimens, analysis of polymorphisms at position 70 of HCV 1b core will remain relevant to identify patients at higher risk of HCC development who require immediate treatment.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Authors' Contributions

V. S. Kichatova and K. K. Kyuregyan contributed equally to this work.

Acknowledgments

This study was supported by the grant of the Russian Science Foundation (Project ID 15-15-30039). Mobility and training of the researchers was supported by the grants of the Swedish Institute (TP 09272/2013 and PI 19806/2016).

Supplementary Materials

Supplementary 1. Figure S1: geographical origins of samples of hepatitis C patients infected with HCV 1b from the Russian Federation used in the analysis of frequency of substitutions in HCV core amino acid positions 70 and 91. Data on the sample set is presented in Table 2.

Supplementary 2. Figure S2: phylogenetic tree for the complete HCV core sequences covering nucleotide positions 342–915 according to HCV 1a reference strain H77 (GenBank accession number NC004102) built using a Bayesian likelihood-based algorithm; detailed overview of the tree fragments (A; B). Sequences of the Russian origins are underlined. Sequences with wild type aa in positions 70 and 91 (R70, L91) are colored green; with mutated R70Q/H, orange; with M91L substitution, blue; with substitutions in both 70 and 91(R70Q/H, M91L), red.

Supplementary 3. Figure S3: T-cell epitopes in the consensus amino acid sequence of nucleocapsid (core) of HCV 1b encompassing amino acids 70 (A) and 91 (B) localized using Immune Epitope Database and Analysis resource (http://www.iedb.org).

Supplementary 4. Table S1: presentation by HLA I alleles of 9-mer peptide variants derived from the region between aa 62 and 78 of the nucleocapsid (core) of HCV 1b encompassing aa 70, predicted using EPISPOT tool (http://bio.med.ucm.es/episopt.html; EPISPOT [25]). Nine amino acid long peptides were chosen based on the consensus amino acid sequence of nucleocapsid (core) protein of HCV 1b variants isolated in the Russian Federation. Upper row: position of amino acids in HCV core; % ($n = 189$), frequency of occurrence of the respective variant in a set of 189 HCV 1b sequences; potential epitopes, selection of potential CTL-epitopes localized within the peptide seqeunce; PPC, population protection coverage (PPC) calculated as the proportion of the population that can potentially mount an immune response to the peptide computed using known frequences of the HLA I alleles that can present the peptide; HLA class I binding profile lists HLA class I molecules predicted to present the peptide. NF, not found.

Supplementary 5. Table S2: presentation by HLA I alleles of 9-mer peptide variants derived from the region between aa 83 and 99 of the nucleocapsid (core) of HCV 1b encompassing aa 91, predicted using EPISPOT tool (http://bio.med.ucm.es/episopt.html; EPISPOT [45]). Nine amino acid long peptides were chosen based on the consensus amino acid sequence of nucleocapsid (core) protein of HCV 1b variants isolated in the Russian Federation. Upper row: position of amino acids in HCV core; % ($n = 189$), percentage of respective variant in a set of 189 HCV 1b sequences; PPC, population protection coverage (PPC) calculated as the proportion of the population that can potentially mount an immune response to the peptide computed using known frequences of the HLA I alleles that can present the epitope; HLA I binding profile lists HLA I molecules predicted to present the epitope. NF, not found.

Supplementary 6. Table S3: NetMHCIIpan 3.1 engine [48] predicts only weak binding of HCV 1b core peptides encompassing amino acid residue 91 to a limited number of human HLA class II molecules. Peptide predicted to be recognized is boxed in green and aa position 91 in red. Strong binding is characterized by binding level < 2%; weak binding, 2 to 10%; no binding, >10%.

References

[1] T. J. Liang, B. Rehermann, L. B. Seeff, and J. H. Hoofnagle, "Pathogenesis, natural history, treatment, and prevention of hepatitis C," *Annals of Internal Medicine*, vol. 132, no. 4, pp. 296–305, 2000.

[2] C. W. Shepard, L. Finelli, and M. J. Alter, "Global epidemiology of hepatitis C virus infection," *The Lancet Infectious Diseases*, vol. 5, no. 9, pp. 558–567, 2005.

[3] J. Thompson Coon, G. Rogers, P. Hewson et al., "Surveillance of cirrhosis for hepatocellular carcinoma: systematic review and economic analysis," *Health Technology Assessment*, vol. 11, no. 34, pp. 1–206, 2007.

[4] A. Petruzziello, S. Marigliano, G. Loquercio, A. Cozzolino, and C. Cacciapuoti, "Global epidemiology of hepatitis C virus infection: an up-date of the distribution and circulation of hepatitis C virus genotypes," *World Journal of Gastroenterology*, vol. 22, no. 34, pp. 7824–7840, 2016.

[5] International Agency for Research on Cancer, *World Cancer Report 2014*, World Health Organization, Geneva, 2014.

[6] G. L. Armstrong, A. Wasley, E. P. Simard, G. M. McQuillan, W. L. Kuhnert, and M. J. Alter, "The prevalence of hepatitis C virus infection in the United States, 1999 through 2002," *Annals of Internal Medicine*, vol. 144, no. 10, pp. 705–714, 2006.

[7] P. Amoroso, M. Rapicetta, M. E. Tosti et al., "Correlation between virus genotype and chronicity rate in acute hepatitis C," *Journal of Hepatology*, vol. 28, no. 6, pp. 939–944, 1998, 9672167.

[8] C. Bréchot, "Hepatitis C virus: molecular biology and genetic variability," *Digestive Diseases and Sciences*, vol. 41, Supplement 12, pp. 6S–21S, 1996.

[9] T. Poynard, P. Bedossa, P. Opolon, and The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC Groups, "Natural history of liver fibrosis progression in patients with chronic hepatitis C," *The Lancet*, vol. 349, no. 9055, pp. 825–832, 1997.

[10] B. Le Guen, G. Squadrito, B. Nalpas, P. Berthelot, S. Pol, and C. Brechot, "Hepatitis C virus genome complexity correlates with response to interferon therapy: a study in French patients with chronic hepatitis C," *Hepatology*, vol. 25, no. 5, pp. 1250–1254, 1997.

[11] J. B. Nousbaum, S. Pol, B. Nalpas et al., "Hepatitis C virus type 1b (II) infection in France and Italy," *Annals of Internal Medicine*, vol. 122, no. 3, pp. 161–168, 1995.

[12] S. Raimondi, S. Bruno, M. U. Mondelli, and P. Maisonneuve, "Hepatitis C virus genotype 1b as a risk factor for hepatocellular carcinoma development: a meta-analysis," *Journal of Hepatology*, vol. 50, no. 6, pp. 1142–1154, 2009.

[13] M. W. Fried, M. L. Shiffman, K. R. Reddy et al., "Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection," *The New England Journal of Medicine*, vol. 347, no. 13, pp. 975–982, 2002.

[14] S. J. Hadziyannis, H. Sette Jr., T. R. Morgan et al., "Peginterferon-α2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose," *Annals of Internal Medicine*, vol. 140, no. 5, pp. 346–355, 2004.

[15] M. P. Manns, J. G. McHutchison, S. C. Gordon et al., "Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial," *The Lancet*, vol. 358, no. 9286, pp. 958–965, 2001.

[16] A. Andriulli, F. Morisco, A. M. Ippolito et al., "HCV genotype 1 subtypes (1a and 1b): similarities and differences in clinical features and therapeutic outcome," *Hepatology International*, vol. 9, no. 1, pp. 52–57, 2015.

[17] R. H. Westbrook and G. Dusheiko, "Natural history of hepatitis C," *Journal of Hepatology*, vol. 61, no. 1, pp. S58–S68, 2014.

[18] N. Akuta, F. Suzuki, H. Sezaki et al., "Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon-ribavirin combination therapy," *Intervirology*, vol. 48, no. 6, pp. 372–380, 2005.

[19] A. El-Shamy and H. Hotta, "Impact of hepatitis C virus heterogeneity on interferon sensitivity: an overview," *World Journal of Gastroenterology*, vol. 20, no. 24, pp. 7555–7569, 2014.

[20] N. Akuta, F. Suzuki, M. Hirakawa et al., "Amino acid substitution in hepatitis C virus core region and genetic variation near the interleukin 28B gene predict viral response to telaprevir with peginterferon and ribavirin," *Hepatology*, vol. 52, no. 2, pp. 421–429, 2010.

[21] N. Akuta, F. Suzuki, Y. Seko et al., "Determinants of response to triple therapy of telaprevir, peginterferon, and ribavirin in previous non-responders infected with HCV genotype 1," *Journal of Medical Virology*, vol. 84, no. 7, pp. 1097–1105, 2012.

[22] N. Akuta, F. Suzuki, Y. Kawamura et al., "Amino acid substitutions in the hepatitis C virus core region are the important predictor of hepatocarcinogenesis," *Hepatology*, vol. 46, no. 5, pp. 1357–1364, 2007.

[23] T. Okanoue, Y. Itoh, H. Hashimoto et al., "Predictive values of amino acid sequences of the core and NS5A regions in antiviral therapy for hepatitis C: a Japanese multi-center study," *Journal of Gastroenterology*, vol. 44, no. 9, pp. 952–963, 2009.

[24] R. C. Jaspe, Y. F. Sulbarán, M. Z. Sulbarán, C. L. Loureiro, H. R. Rangel, and F. H. Pujol, "Prevalence of amino acid mutations in hepatitis C virus core and NS5B regions among Venezuelan viral isolates and comparison with worldwide isolates," *Virology Journal*, vol. 9, no. 1, p. 214, 2012.

[25] C. Sultana, G. Oprişan, M. D. Teleman et al., "Impact of hepatitis C virus core mutations on the response to interferon-based treatment in chronic hepatitis C," *World Journal of Gastroenterology*, vol. 22, no. 37, pp. 8406–8413, 2016.

[26] E. Alestig, B. Arnholm, A. Eilard et al., "Core mutations, IL28B polymorphisms and response to peginterferon/ribavirin treatment in Swedish patients with hepatitis C virus genotype 1 infection," *BMC Infectious Diseases*, vol. 11, no. 1, article 124, 2011.

[27] M. J. Donlin, N. A. Cannon, E. Yao et al., "Pretreatment sequence diversity differences in the full-length hepatitis C virus open reading frame correlate with early response to therapy," *Journal of Virology*, vol. 81, no. 15, pp. 8211–8224, 2007.

[28] F. A. Di Lello, J. A. Mira, K. Neukam et al., "Core amino acid variation at position 110 is associated with sustained virological response in Caucasian patients with chronic hepatitis C virus 1b infection," *Archives of Virology*, vol. 159, no. 12, pp. 3345–3351, 2014.

[29] F. S. Alhamlan, M. N. Al-Ahdal, N. Z. Khalaf et al., "Hepatitis C virus genotype 1: how genetic variability of the core protein affects the response to pegylated-interferon and ribavirin therapy," *Journal of Medical Virology*, vol. 86, no. 2, pp. 224–234, 2014.

[30] M. Tasaka-Fujita, N. Sugiyama, W. Kang et al., "Amino acid polymorphisms in hepatitis C virus core affect infectious virus production and major histocompatibility complex class I molecule expression," *Scientific Reports*, vol. 5, article 13994, 2015.

[31] A. El-Shamy, M. Pendleton, F. J. Eng, E. H. Doyle, A. Bashir, and A. D. Branch, "Impact of HCV core gene quasispecies on hepatocellular carcinoma risk among HALT-C trial patients," *Scientific Reports*, vol. 6, article 27025, 2016.

[32] N. Akuta, F. Suzuki, Y. Seko et al., "Complicated relationships of amino acid substitution in hepatitis C virus core region and *IL28B* genotype influencing hepatocarcinogenesis," *Hepatology*, vol. 56, no. 6, pp. 2134–2141, 2012.

[33] S. Nakamoto, F. Imazeki, K. Fukai et al., "Association between mutations in the core region of hepatitis C virus genotype 1 and hepatocellular carcinoma development," *Journal of Hepatology*, vol. 52, no. 1, pp. 72–78, 2010.

[34] N. Akuta, F. Suzuki, M. Hirakawa et al., "Amino acid substitutions in hepatitis C virus core region predict hepatocarcinogenesis following eradication of HCV RNA by antiviral therapy," *Journal of Medical Virology*, vol. 83, no. 6, pp. 1016–1022, 2011.

[35] Y. Seko, N. Akuta, F. Suzuki et al., "Amino acid substitutions in the hepatitis C virus core region and lipid metabolism are associated with hepatocarcinogenesis in nonresponders to interferon plus ribavirin combination therapy," *Intervirology*, vol. 56, no. 1, pp. 13–21, 2013.

[36] A. El-Shamy, M. Shindo, I. Shoji, L. Deng, T. Okuno, and H. Hotta, "Polymorphisms of the core, NS3, and NS5A proteins of hepatitis C virus genotype 1b associate with development of hepatocellular carcinoma," *Hepatology*, vol. 58, no. 2, pp. 555–563, 2013.

[37] N. Akuta, F. Suzuki, M. Kobayashi et al., "Impact of mutations at amino acid 70 in hepatitis C virus (HCV) genotype 1b core region on hepatocarcinogenesis following eradication of HCV

RNA," *Journal of Clinical Microbiology*, vol. 53, no. 9, pp. 3039–3041, 2015.

[38] S. Kobayashi, T. Takeda, M. Enomoto et al., "Development of hepatocellular carcinoma in patients with chronic hepatitis C who had a sustained virological response to interferon therapy: a multicenter, retrospective cohort study of 1124 patients," *Liver International*, vol. 27, no. 2, pp. 186–191, 2007.

[39] H. Bailey, A. Turkova, and C. Thorne, "Syphilis, hepatitis C and HIV in Eastern Europe," *Current Opinion in Infectious Diseases*, vol. 30, no. 1, pp. 93–100, 2017.

[40] A. V. Shustov, G. V. Kochneva, G. F. Sivolobova et al., "Molecular epidemiology of the hepatitis C virus in Western Siberia," *Journal of Medical Virology*, vol. 77, no. 3, pp. 382–389, 2005.

[41] S. N. Kuzin, E. I. Samokhvalov, E. E. Zabotina et al., "Hepatitis virus genotype structure in patients with chronic hepatitis C," *Zhurnal Mikrobiologii, Epidemiologii, i Immunobiologii*, no. 3, pp. 33–38, 2011.

[42] V. Kartashev, M. Döring, L. Nieto et al., "New findings in HCV genotype distribution in selected West European, Russian and Israeli regions," *Journal of Clinical Virology*, vol. 81, pp. 82–89, 2016.

[43] A. J. Drummond and A. Rambaut, "BEAST: Bayesian evolutionary analysis by sampling trees," *BMC Evolutionary Biology*, vol. 7, no. 1, p. 214, 2007.

[44] R. Vita, J. A. Overton, J. A. Greenbaum et al., "The immune epitope database (IEDB) 3.0," *Nucleic Acids Research*, vol. 43, no. D1, pp. D405–D412, 2014.

[45] M. Molero-Abraham, E. M. Lafuente, D. R. Flower, and P. A. Reche, "Selection of conserved epitopes from hepatitis C virus for pan-populational stimulation of T-cell responses," *Clinical and Developmental Immunology*, vol. 2013, Article ID 601943, 10 pages, 2013.

[46] L. Lebedeva, A. Chumak, T. Pukhlikova, S. Pavlenko, V. Zinkin, and O. Mayorova, "The features of histocompatibility antigens polymorphism of cord blood samples in Moscow Blood Bank," *Bulletin of Hematology*, vol. 11, no. 2, pp. 20-21, 2015.

[47] F. F. González-Galarza, L. Y. C. Takeshita, E. J. M. Santos et al., "Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations," *Nucleic Acid Research*, vol. 43, no. D1, pp. D784–D788, 2015.

[48] M. Andreatta, E. Karosiene, M. Rasmussen, A. Stryhn, S. Buus, and M. Nielsen, "Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification," *Immunogenetics*, vol. 67, no. 11-12, pp. 641–650, 2015.

[49] I. Hoof, B. Peters, J. Sidney et al., "NetMHCpan, a method for MHC class I binding prediction beyond humans," *Immunogenetics*, vol. 61, no. 1, pp. 1–13, 2009.

[50] M. Nielsen and M. Andreatta, "NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets," *Genome Medicine*, vol. 8, p. 33, 2016.

[51] A. K. Akhmineeva, *HLA Phenotype and Clinical Course of Chronic Viral Hepatitis, [Ph.D. thesis]*, Astrakhan State Medical Academy, Astrakhan, 2008.

[52] A. Utama, N. P. Tania, R. Dhenni et al., "Genotype diversity of hepatitis C virus (HCV) in HCV-associated liver disease patients in Indonesia," *Liver International*, vol. 30, no. 8, pp. 1152–1160, 2010.

[53] I. Brahim, S. Ezzikouri, E. M. Mtairag et al., "Amino acid substitutions in the hepatitis C virus core region of genotype 1b in Moroccan patients," *Infection, Genetics and Evolution*, vol. 14, pp. 102–104, 2013.

[54] S. C. Ray, L. Fanning, X. H. Wang, D. M. Netski, E. Kenny-Walsh, and D. L. Thomas, "Divergent and convergent evolution after a common-source outbreak of hepatitis C virus," *Journal of Experimental Medicine*, vol. 201, no. 11, pp. 1753–1759, 2005.

[55] R. Aurora, M. J. Donlin, N. A. Cannon, and J. E. Tavis, "Genome-wide hepatitis C virus amino acid covariance networks can predict response to antiviral therapy in humans," *The Journal of Clinical Investigation*, vol. 119, no. 1, pp. 225–236, 2008.

[56] Z. Hu, Y. Liu, L. Qiu et al., "Kinetic response of wild and mutant core codon 70 strains of HCV genotype 1b to pegylated interferon-α and ribavirin therapy," *Virology Journal*, vol. 12, no. 1, p. 220, 2015.

[57] K. R. Reddy, J. H. Hoofnagle, M. J. Tong et al., "Racial differences in responses to therapy with interferon in chronic hepatitis C," *Hepatology*, vol. 30, no. 3, pp. 787–793, 1999.

[58] M. Isaguliants and N. Ozeretskovskaya, "Host background factors contributing to hepatitis C virus clearance," *Current Pharmaceutical Biotechnology*, vol. 4, no. 3, pp. 185–193, 2003.

[59] Y. Tanaka, N. Nishida, M. Sugiyama et al., "Genome-wide association of *IL28B* with response to pegylated interferon-α and ribavirin therapy for chronic hepatitis C," *Nature Genetics*, vol. 41, no. 10, pp. 1105–1109, 2009.

[60] C. Fabris, E. Falleti, A. Cussigh et al., "IL-28B rs12979860 C/T allele distribution in patients with liver cirrhosis: role in the course of chronic viral hepatitis and the development of HCC," *Journal of Hepatology*, vol. 54, no. 4, pp. 716–722, 2011.

[61] Y. Funaoka, N. Sakamoto, G. Suda et al., "Analysis of interferon signaling by infectious hepatitis C virus clones with substitutions of core amino acids 70 and 91," *Journal of Virology*, vol. 85, no. 12, pp. 5986–5994, 2011.

[62] N. Akuta, F. Suzuki, M. Hirakawa et al., "Amino acid substitutions in the hepatitis C virus core region of genotype 1b are the important predictor of severe insulin resistance in patients without cirrhosis and diabetes mellitus," *Journal of Medical Virology*, vol. 81, no. 6, pp. 1032–1039, 2009.

[63] S. Uraki, M. Tameda, K. Sugimoto et al., "Substitution in amino acid 70 of hepatitis C virus core protein changes the adipokine profile via toll-like receptor 2/4 signaling," *PLoS One*, vol. 10, no. 6, article e0131346, 2015.

[64] Y. Tachi, Y. Katano, T. Honda et al., "Impact of amino acid substitutions in the hepatitis C virus genotype 1b core region on liver steatosis and hepatic oxidative stress in patients with chronic hepatitis C," *Liver International*, vol. 30, no. 4, pp. 554–559, 2010.

[65] Y. Sumida, K. Kanemasa, T. Hara et al., "Impact of amino acid substitutions in hepatitis C virus genotype 1b core region on liver steatosis and glucose tolerance in non-cirrhotic patients without overt diabetes," *Journal of Gastroenterology and Hepatology*, vol. 26, no. 5, pp. 836–842, 2011.

[66] A. V. Ivanov, V. T. Valuev-Elliston, D. A. Tyurina et al., "Oxidative stress, a trigger of hepatitis C and B virus-induced liver carcinogenesis," *Oncotarget*, vol. 8, no. 3, pp. 3895–3932, 2017.

[67] M. J. Donlin, N. A. Cannon, R. Aurora et al., "Contribution of genome-wide HCV genetic differences to outcome of interferon-based therapy in Caucasian American and

African American patients," *PLoS One*, vol. 5, no. 2, article e9032, 2010.

[68] L. N. Cavalcante, K. Abe-Sandes, A. L. D. Angelo et al., "*IL28B* polymorphisms are markers of therapy response and are influenced by genetic ancestry in chronic hepatitis C patients from an admixed population," *Liver International*, vol. 32, no. 3, pp. 476–486, 2012.

[69] U. Navaneethan, N. Kemmer, and G. W. Neff, "Review: predicting the probable outcome of treatment in HCV patients," *Therapeutic Advances in Gastroenterology*, vol. 2, no. 5, pp. 287–302, 2009.

[70] J. E. Layden-Almer, R. M. Ribeiro, T. Wiley, A. S. Perelson, and T. J. Layden, "Viral dynamics and response differences in HCV-infected African American and white patients treated with IFN and ribavirin," *Hepatology*, vol. 37, no. 6, pp. 1343–1350, 2003.

[71] J. H. Kao, S. H. Ahn, R. N. Chien et al., "Urgency to treat patients with chronic hepatitis C in Asia," *Journal of Gastroenterology and Hepatology*, vol. 32, no. 5, pp. 966–974, 2017.

[72] N. Enomoto, I. Sakuma, Y. Asahina et al., "Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region," *The Journal of Clinical Investigation*, vol. 96, no. 1, pp. 224–230, 1995.

[73] M. Ragonnet-Cronin, S. Aris-Brosou, I. Joanisse et al., "Adaptive evolution of HIV at HLA epitopes is associated with ethnicity in Canada," *PLoS One*, vol. 7, no. 5, article e36933, 2012.

[74] J. Á. Patiño-Galindo and F. González-Candelas, "Comparative analysis of variation and selection in the HCV genome," *Infection, Genetics and Evolution*, vol. 49, pp. 104–110, 2017.

[75] S. L. Rhodes, H. Erlich, K. A. Im et al., "Associations between the human MHC and sustained virologic response in the treatment of chronic hepatitis C virus infection," *Genes & Immunity*, vol. 9, no. 4, pp. 328–333, 2008.

Hindawi
Journal of Immunology Research
Volume 2017, Article ID 2680160, 14 pages
https://doi.org/10.1155/2017/2680160

Review Article

Fundamentals and Methods for T- and B-Cell Epitope Prediction

Jose L. Sanchez-Trincado, Marta Gomez-Perosanz, and Pedro A. Reche

Laboratory of Immunomedicine, Faculty of Medicine, Complutense University of Madrid, Ave Complutense S/N, 28040 Madrid, Spain

Correspondence should be addressed to Pedro A. Reche; parecheg@med.ucm.es

Received 27 July 2017; Revised 22 November 2017; Accepted 27 November 2017; Published 28 December 2017

Academic Editor: Senthami R. Selvan

Adaptive immunity is mediated by T- and B-cells, which are immune cells capable of developing pathogen-specific memory that confers immunological protection. Memory and effector functions of B- and T-cells are predicated on the recognition through specialized receptors of specific targets (antigens) in pathogens. More specifically, B- and T-cells recognize portions within their cognate antigens known as epitopes. There is great interest in identifying epitopes in antigens for a number of practical reasons, including understanding disease etiology, immune monitoring, developing diagnosis assays, and designing epitope-based vaccines. Epitope identification is costly and time-consuming as it requires experimental screening of large arrays of potential epitope candidates. Fortunately, researchers have developed in silico prediction methods that dramatically reduce the burden associated with epitope mapping by decreasing the list of potential epitope candidates for experimental testing. Here, we analyze aspects of antigen recognition by T- and B-cells that are relevant for epitope prediction. Subsequently, we provide a systematic and inclusive review of the most relevant B- and T-cell epitope prediction methods and tools, paying particular attention to their foundations.

1. Introduction

The immune system is typically divided into two categories, innate and adaptive. Innate immunity involves nonspecific defense mechanisms that act immediately or within hours after a microbe appearance in the body. All multicellular beings exhibit some kind of innate immunity. In contrast, adaptive immunity is only present in vertebrates and it is highly specific. In fact, the adaptive immune system is able to recognize and destroy invading pathogens individually. Moreover, the adaptive immune system remembers the pathogens that fights, acquiring a pathogen-specific long-lasting protective memory that enables stronger attacks each time the pathogen is reencountered [1]. Nonetheless, innate and adaptive immune mechanisms work together and adaptive immunity elicitation is contingent on prior activation of innate immune responses [1].

Adaptive immunity is articulated by lymphocytes, more specifically by B- and T-cells, which are responsible for the humoral and cell-mediated immunity. B- and T-cells do not recognize pathogens as a whole, but molecular components known as antigens. These antigens are recognized by specific receptors present in the cell surface of B- and T-cells. Antigen recognition by these receptors is required to activate B- and T-cells but not enough, as second activation signals stemming from the activation of the innate immune system are also needed. The specificity of the recognition is determined by genetic recombination events that occur during lymphocyte development, which lead to generating millions of different variants of lymphocytes in terms of the antigen-recognizing receptors [1]. Antigen recognition by B- and T-cells differ greatly.

B-cells recognize solvent-exposed antigens through antigen receptors, named as B-cell receptors (BCR), consisting of membrane-bound immunoglobulins, as shown in Figure 1. Upon activation, B-cells differentiate and secrete soluble forms of the immunoglobulins, also known as antibodies, which mediate humoral adaptive immunity.

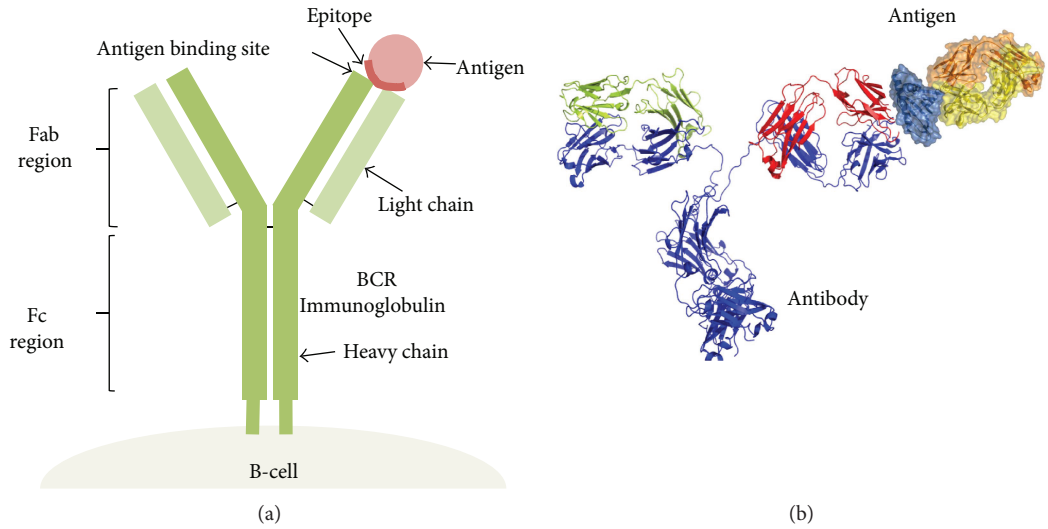


Figure 1: B-cell epitope recognition. B-cell epitopes are solvent-exposed portions of the antigen that bind to secreted and cell-bound immunoglobulins. (a) B-cell receptors encompass cell-bound immunoglobulins, consisting of two heavy chains and two light chains. The different chains and regions are annotated. (b) Molecular representation of the interaction between an antibody and the antigen. Antibodies are secreted immunoglobulins of known specificity.

Antibodies released by B-cells can have different functions that are triggered upon binding their cognate antigens. These functions include neutralizing toxins and pathogens and labeling them for destruction [1].

A B-cell epitope is the antigen portion binding to the immunoglobulin or antibody. These epitopes recognized by B-cells may constitute any exposed solvent region in the antigen and can be of different chemical nature. However, most antigens are proteins and those are the subjects for epitope prediction methods.

On the other hand, T-cells present on their surface a specific receptor known as T-cell receptor (TCR) that enables the recognition of antigens when they are displayed on the surface of antigen-presenting cells (APCs) bound to major histocompatibility complex (MHC) molecules. T-cell epitopes are presented by class I (MHC I) and II (MHC II) MHC molecules that are recognized by two distinct subsets of T-cells, CD8 and CD4 T-cells, respectively (Figure 2). Subsequently, there are CD8 and CD4 T-cell epitopes. CD8 T-cells become cytotoxic T lymphocytes (CTL) following T CD8 epitope recognition. Meanwhile, primed CD4 T-cells become helper (Th) or regulatory (Treg) T-cells [1]. Th cells amplify the immune response, and there are three main subclasses: Th1 (cell-mediated immunity against intracellular pathogens), Th2 (antibody-mediated immunity), and Th17 (inflammatory response and defense against extracellular bacteria) [2].

Identifying epitopes in antigens is of great interest for a number of practical reasons, including understanding disease etiology, immune monitoring, developing diagnosis assays, and designing epitope-based vaccines. B-cell epitopes can be identified by different methods including solving the 3D structure of antigen-antibody complexes, peptide library screening of antibody binding or performing functional assays in which the antigen is mutated and the interaction antibody-antigen is evaluated [3, 4]. On the other hand, experimental determination of T-cell epitopes is carried out using MHC multimers and lymphoproliferation or ELISPOT assays, among others [5, 6]. Traditional epitope identification has depended entirely upon experimental techniques, being costly and time-consuming. Thereby, scientists have developed and implemented epitope prediction methods that facilitate epitope identification and decrease the experimental load associated with it. Here, we will first analyze aspects of antigen recognition by T- and B-cells that are relevant for a better understanding of the topic of epitope prediction. Subsequently, we will provide a systematic and inclusive review of the most important prediction methods and tools, paying particular attention to their foundations and potentials. We will also discuss epitope prediction limitations and ways to overcome them. We will start with T-cell epitopes.

2. T-Cell Epitope Prediction

T-cell epitope prediction aims to identify the shortest peptides within an antigen that are able to stimulate either CD4 or CD8 T-cells [7]. This capacity to stimulate T-cells is called immunogenicity, and it is confirmed in assays requiring synthetic peptides derived from antigens [5, 6]. There are many distinct peptides within antigens and T-cell prediction methods aim to identify those that are immunogenic. T-cell epitope immunogenicity is contingent on three basic steps: (i) antigen processing, (ii) peptide binding to MHC molecules, and (iii) recognition by a cognate TCR. Of these three events, MHC-peptide binding is the most selective one at determining T-cell epitopes [8, 9]. Therefore, prediction of peptide-MHC binding is the main basis to anticipate T-cell epitopes and we will review it next.

2.1. Prediction of Peptide-MHC Binding. MHC I and MHC II molecules have similar 3D-structures with bound peptides sitting in a groove delineated by two α-helices overlying a

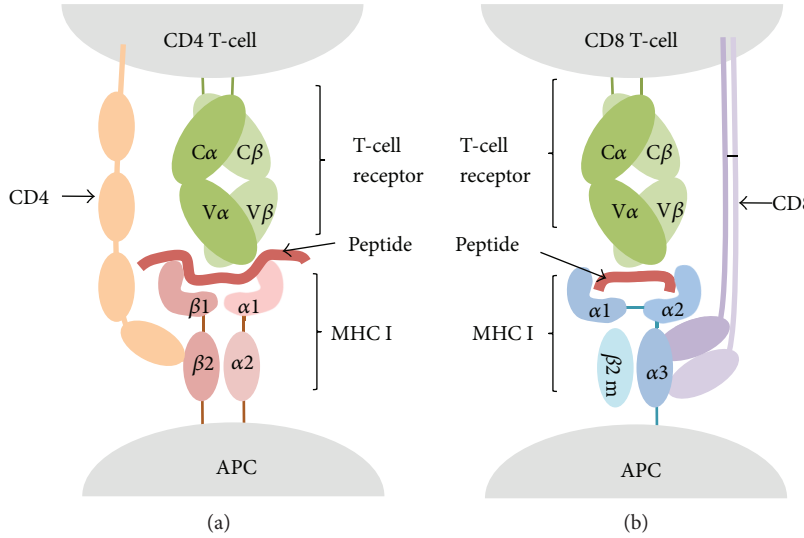


Figure 2: T-cell epitope recognition. T-cell epitopes are peptides derived from antigens and recognized by the T-cell receptor (TCR) when bound to MHC molecules displayed on the cell surface of APCs. (a) CD4 T-cells express the CD4 coreceptor, which binds to MHC II, and recognize peptides presented by MHC II molecules. (b) CD8 T-cells express the CD8 coreceptor, which binds to MHC I, and recognize peptides presented by MHC I molecules.

floor comprised of eight antiparallel β-strands. However, there are also key differences between MHC I and II binding grooves that we must highlight for they condition peptide-binding predictions (Figure 3). The peptide-binding cleft of MHC I molecules is closed as it is made by a single α chain. As a result, MHC I molecules can only bind short peptides ranging from 9 to 11 amino acids, whose N- and C-terminal ends remain pinned to conserved residues of the MHC I molecule through a network of hydrogen bonds [10, 11]. The MHC I peptide-binding groove also contains deep binding pockets with tight physicochemical preferences that facilitate binding predictions. There is a complication however. Peptides that have different sizes and bind to the same MHC I molecule often use alternative binding pockets [12]. Therefore, methods predicting peptide-MHC I binding require a fixed peptide length. However, since most MHC I peptide ligands have 9 residues, it is generally preferable to predict peptides with that size. In contrast, the peptide-binding groove of MHC II molecules is open, allowing the N- and C-terminal ends of a peptide to extend beyond the binding groove [10, 11]. As a result, MHC II-bound peptides vary widely in length (9–22 residues), although only a core of nine residues (peptide-binding core) sits into the MHC II binding groove. Therefore, peptide-MHC II binding prediction methods often target to identify these peptide-binding cores. MHC II molecule binding pockets are also shallower and less demanding than those of MHC I molecules. As a consequence, peptide-binding prediction to MHC II molecules is less accurate than that of MHC I molecules.

Given the relevance of the problem, there are numerous methods to predict peptide-MHC binding. The most relevant with free online use are collected on Table 1. They can be divided in two main categories: data-driven and structure-based methods. Structure-based approaches generally rely on modeling the peptide-MHC structure followed by evaluation of the interaction through methods such as molecular dynamic simulations [8, 13]. Structure-based methods have the great advantage of not needing experimental data. However, they are seldom used as they are computationally intensive and exhibit lower predictive performance than data-driven methods [14].

Data-driven methods for peptide-MHC binding prediction are based on peptide sequences that are known to bind to MHC molecules. These peptide sequences are generally available in specialized epitope databases such as IEDB [15], EPIMHC [16], Antijen [17, 18]. Both MHC I and II binding peptides contain frequently occurring amino acids at particular peptide positions, known as anchor residues. Thereby, prediction of peptide-MHC binding was first approached using sequence motif (SM) reflecting amino acid preferences of MHC molecules at anchor positions [19]. However, it was soon shown that nonanchor residues also contribute to the capacity of a peptide to bind to a given MHC molecule [20, 21]. Subsequently, researchers developed motif matrices (MM), which could evaluate the contribution of each and all peptide positions to the binding with the MHC molecule [22–25]. The most sophisticated form of motif matrices consists of profiles [24–26] that are similar to those used for detecting sequence homology [27]. We would like to remark that motif matrices are often mistaken with quantitative affinity matrices (QAMs) since both produce peptide scores. However, MMs are derived without taking in consideration values of binding affinities and, therefore, resulting peptide scores are not suited to address binding affinity. In contrast, QAMs are trained on peptides and corresponding binding affinities, and aim to predict binding affinity. The first method based on QAMs was developed by Parker et al. [28] (Table 1). Subsequently, various approaches were developed to obtain QAMs from peptide

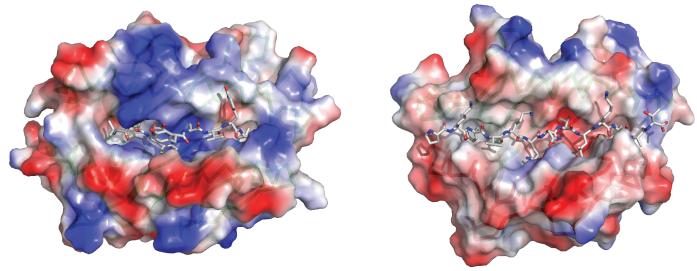


Figure 3: MHC molecule binding groove. The figure depicts the molecular surface as seen by the TCR of representative MHC I and II molecules. Note how the binding groove of the MHC I molecule is closed but that of MHC II is open. As a result, MHC I molecules bind short peptides (8–11 amino acids), while MHC II molecules bind longer peptides (9–22 amino acids). The figure was prepared from PDB files 1QRN (MHC I) and 1FYT (MHC II) using PyMol.

Table 1: Selected T-cell epitope prediction tools available online for free public use.

Tool	URL	Method[1]	MHC class	A	S	T	P	Ref.
EpiDOCK	http://epidock.ddg-pharmfac.net	SB	II	—	—	—	—	[86]
MotifScan	https://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan	SM	I and II	—	X	—	—	—
Rankpep	http://imed.med.ucm.es/Tools/rankpep.html	MM	I and II	—	—	—	X	[26]
SYFPEITHI	http://www.syfpeithi.de/	MM	I and II	—	—	—	—	[23]
MAPPP	http://www.mpiib-berlin.mpg.de/MAPPP/	MM	I	—	X	—	X	[87]
PREDIVAC	http://predivac.biosci.uq.edu.au/	MM	II	—	—	—	—	[88]
PEPVAC	http://imed.med.ucm.es/PEPVAC/	MM	I	—	X	—	X	[63]
EPISOPT	http://bio.med.ucm.es/episopt.html	MM	I	—	X	—	—	[64]
Vaxign	http://www.violinet.org/vaxign/	MM	I and II	—	—	—	—	[89]
MHCPred	http://www.ddg-pharmfac.net/mhcpred/MHCPred/	QSAR	I and II	X	—	—	—	[34]
EpiTOP	http://www.pharmfac.net/EpiTOP	QSAR	II	X	—	—	—	[90]
BIMAS	https://www-bimas.cit.nih.gov/molbio/hla_bind/	QAM	I	X				[28]
TEPITOPE	http://datamining-iip.fudan.edu.cn/service/TEPITOPEpan/TEPITOPEpan.html	QAM	II	X	—	—	—	[32]
Propred	http://www.imtech.res.in/raghava/propred/	QAM	II	X	X	—	—	[91]
Propred-1	http://www.imtech.res.in/raghava/propred1/	QAM	I	X	X	—	X	[92]
EpiJen	http://www.ddg-pharmfac.net/epijen/EpiJen/EpiJen.htm	QAM	I	X	—	X	X	[82]
IEDB-MHCI	http://tools.immuneepitope.org/mhci/	Combined	I	X	—	—	—	[93]
IEDB-MHCII	http://tools.immuneepitope.org/mhcii/	Combined	II	X	—	—	—	[93]
IL4pred	http://webs.iiitd.edu.in/raghava/il4pred/index.php	SVM	II	—	—	—	—	[67]
MULTIPRED2	http://cvc.dfci.harvard.edu/multipred2/index.php	ANN	I and II	—	X	—	—	[62]
MHC2PRED	http://www.imtech.res.in/raghava/mhc2pred/index.html	SVM	II	—	—	—	—	[38]
NetMHC	http://www.cbs.dtu.dk/services/NetMHC/	ANN	I	X	—	—	—	[49]
NetMHCII	http://www.cbs.dtu.dk/services/NetMHCII/	ANN	II	X	—	—	—	[30]
NetMHCpan	http://www.cbs.dtu.dk/services/NetMHCpan/	ANN	I	X	—	—	—	[54]
NetMHCIIpan	http://www.cbs.dtu.dk/services/NetMHCIIpan/	ANN	II	X	—	—	—	[55]
nHLApred	http://www.imtech.res.in/raghava/nhlapred/	ANN	I	—	—	—	X	[94]
SVMHC	http://abi.inf.uni-tuebingen.de/Services/SVMHC/	SVM	I and II	—	—	—	—	[95]
SVRMHC	http://us.accurascience.com/SVRMHCdb/	SVM	I and II	X	—	—	—	[46]
NetCTL	http://www.cbs.dtu.dk/services/NetCTL/	ANN	I	X	X	X	X	[83]
WAPP	https://abi.inf.uni-tuebingen.de/Services/WAPP/index_html	SVM	I	—	—	X	X	[37]

[1]Method used for prediction of peptide-MHC binding. Keys for methods: SM: sequence motif; SB: structure-based; MM: motif matrix; QAM: quantitative affinity matrix; SVM: support vector machine; ANN: artificial neural network; QSAR: quantitative structure-activity relationship model; combined: tool uses different methods including ANN and QAM, selecting the more appropriate method for each distinct MHC molecule. The table also indicates whether the tools predict quantitative binding affinity (A), supertypes (S), TAP binding (T), and proteasomal cleavage (P); marked with an X in the affirmative case.

affinity data and predict peptide binding to MHC I and II molecules [29–32].

QAMs and motif matrices assume an independent contribution of peptide side chains to the binding. This assumption is well supported by experimental data but there is also evidence that neighboring peptide residues interfere with others [33]. To account for those interferences, researchers introduced quantitative structure-activity relationship (QSAR) additive models wherein the binding affinity of peptides to MHC is computed as the sum of amino acid contributions at each position plus the contribution of adjacent side chain interactions [34]. However, machine learning (ML) is the most popular and robust approach introduced to deal with the nonlinearity of peptide-MHC binding data [8]. Researchers have used ML for two distinct problems: the discrimination of MHC binders from nonbinders and the prediction of binding affinity of peptides to MHC molecules.

For developing discrimination models, ML algorithms are trained on data sets consisting of peptides that either bind or do not bind to MHC molecules. Relevant examples of ML-based discrimination models are those based on artificial neural networks (ANNs) [35, 36], support vector machines (SVMs) [37–39], decision trees (DTs) [40, 41], and Hidden Markov models (HMMs), which can also cope with nonlinear data and have been used to discriminate peptides binding to MHC molecules. However, unlike other ML algorithms, they have to be trained only on positive data. Three types of HMMs have been used to predict MHC-peptide binding: fully connected HMMs [42], structure-optimized HMMs [43], and profile HMMs [43, 44]. Of these, only fully connected HMMs (fcHMMs) and structure-optimized HMMs (soHMMs) can recognize different patterns in the peptide binders. In fact, profile HMMs that are derived from sets of ungapped alignments (the case for peptides binding to MHC) are nearly identical to profile matrices [45] (Table 1).

With regard to predicting binding affinity, ML algorithms are trained on datasets consisting of peptides with known affinity to MHC molecules. Both SVMs and ANNs have been used for such purpose. SVMs were first applied to predict peptide-binding affinity to MHC I molecules [46] and later to MHC II molecules [47] (Table 1). Likewise, ANNs were also applied first to the prediction of peptide binding to MHC I [48, 49] and later to MHC II molecules [50] (Table 1). Benchmarking of peptide-MHC binding prediction methods appears to indicate that those based on ANNs are superior to those based on QAMs and MMs. However, the differences between the distinct methods are marginal and vary for different MHC molecules [51]. Moreover, it has been shown that the performance of peptide-MHC predictions is improved by combining several methods and providing consensus predictions [52].

A major complication for predicting T-cell epitopes through peptide-MHC binding models is MHC polymorphism. In humans, MHC molecules are known as human leukocyte antigens (HLAs), and there are hundreds of allelic variants of class I (HLA I) and class II (HLA II) molecules. These HLA allelic variants bind distinct sets of peptides [53] and require specific models for predicting peptide-MHC binding. However, peptide-binding data is only available for a minority of HLA molecules. To overcome this limitation, some researchers have developed pan-MHC-specific methods by training ANNs on input data combining MHC residues that contact the peptide with peptide-binding affinity that are capable of predicting peptide-binding affinities to uncharacterized HLA alleles [54, 55].

HLA polymorphism also hampers the development of worldwide covering T-cell epitope-based vaccines as HLA variants are expressed at vastly variable frequencies in different ethnic groups [56]. Interestingly, different HLA molecules can also bind similar sets of peptides [57, 58] and researchers have devised methods to cluster them in groups, known as HLA supertypes, consisting of HLA alleles with similar peptide-binding specificities [59–61]. The HLA-A2, HLA-A3, and HLA-B7 are relevant examples of supertypes; 88% of the population expresses at least an allele included in these supertypes [25, 57, 58]. Identification of promiscuous peptide-binding to HLA supertypes enables the development of T-cell epitope vaccines with high-population coverage using a limited number of peptides. Currently, several web-based methods allow the prediction of promiscuous peptide-binding to HLA supertypes for epitope vaccine design including MULTIPRED [62] and PEPVAC [63] (Table 1). A method to identify promiscuous peptide-binding beyond HLA supertypes was developed and implemented by Molero-Abraham et al. [64] with the name of EPISOPT. EPISOPT predicts HLA I presentation profiles of individual peptides regardless of supertypes and identifies epitope combinations providing a wider population protection coverage.

Prediction of peptide binding to MHC II molecules readily discriminate CD4 T-cell epitopes, but cannot tell their ability to activate the response of specific CD4 T-cell subsets (e.g., Th1, Th2, and Treg). However, there is evidence that some CD4 T-cell epitopes appear to stimulate specific subsets of Th cells [65, 66]. Distinguishing the ability of MHC II-restricted epitopes to elicit distinct responses is clearly relevant for epitope vaccine development and has prompted researchers' attention. A relevant example is the work by Dhanda et al. [67] who generated classifiers capable of predicting potential peptide inducers of interleukin 4 (IL-4) secretion, typical of Th2 cells, by training SVM models on experimentally validated IL4 inducing and noninducing MHC class II binders (Table 1).

2.2. Prediction of Antigen Processing and Integration with Peptide-MHC Binding Prediction. Antigen processing shapes the peptide repertoire available for MHC binding and is a limiting step determining T-cell epitope immunogenicity [68]. Subsequently, computational modeling of the antigen processing pathway provides a mean to enhance T-cell epitope predictions. Antigen presentation by MHC I and II molecules proceed by two different pathways. MHC II molecules present peptide antigens derived from endocyted antigens that are degraded and loaded onto the MHC II molecule in endosomal compartments [69]. Class II antigen degradation is poorly understood, and there is lack of good prediction algorithms yet [70]. In contrast, MHC I molecules

present peptides derived mainly from antigens degraded in the cytosol. The resulting peptide antigens are then transported to the endoplasmic reticulum by TAP where they are loaded onto nascent MHC I molecules [69] (Figure 4). Prior to loading, peptides often undergo trimming by ERAAP N-terminal amino peptidases [71].

Proteasomal cleavage and peptide-binding to TAP have been studied in detail and there are computational methods that predict both processes. Proteasomal cleavage prediction models have been derived from peptide fragments generated *in vitro* by human constitutive proteasomes [72, 73] and from sets of MHC I-restricted ligands mapped onto their source proteins [74–76]. On the other hand, TAP binding prediction methods have been developed by training different algorithms on peptides of known affinity to TAP [77–80]. Combination of proteasomal cleavage and peptide-binding to TAP with peptide-MHC binding predictions increases T-cell epitope predictive rate in comparison to just peptide-binding to MHC I [37, 77, 81–83]. Subsequently, researchers have developed resources to predict CD8 T-cell epitopes through multistep approaches integrating proteasomal cleavage, TAP transport, and peptide-binding to MHC molecules [26, 37, 82–85] (Table 1).

3. Prediction of B-Cell Epitopes

B-cell epitope prediction aims to facilitate B-cell epitope identification with the practical purpose of replacing the antigen for antibody production or for carrying structure-function studies. Any solvent-exposed region in the antigen can be subject of recognition by antibodies. Nonetheless, B-cell epitopes can be divided in two main groups: linear and conformational (Figure 5). Linear B-cell epitopes consist of sequential residues, peptides, whereas conformational B-cell epitopes consist of patches of solvent-exposed atoms from residues that are not necessarily sequential (Figure 5). Therefore, linear and conformational B-cell epitopes are also known as continuous and discontinuous B-cell epitopes, respectively. Antibodies recognizing linear B-cell epitopes can recognize denatured antigens, while denaturing the antigen results in loss of recognition for conformational B-cell epitopes. Most B-cell epitopes (approximately a 90%) are conformational and, in fact, only a minority of native antigens contains linear B-cell epitopes [3]. We will review both, prediction of linear and conformational B-cell epitopes.

3.1. Prediction of Linear B-Cell Epitopes. Linear B-cell epitopes consist of peptides which can readily be used to replace antigens for immunizations and antibody production. Therefore, despite being a minority, prediction of linear B-cell epitopes have received major attention. Linear B-cell epitopes are predicted from the primary sequence of antigens using sequence-based methods. Early computational methods for the prediction of B-cell epitopes were based on simple amino acid propensity scales depicting physicochemical features of B-cellepitopes. For example, Hopp and Wood applied residue hydrophilicity calculations for B-cell epitope prediction [96, 97] on the assumption that hydrophilic regions are predominantly located on the protein surface and are potentially antigenic. We know now, however, that protein surfaces contain roughly the same number of hydrophilic and hydrophobic residues [98]. Other amino acid propensity scales introduced for B-cell epitope prediction are based on flexibility [99], surface accessibility [100], and β-turn propensity [101]. Current available bioinformatics tools to predict linear B-cell epitopes using propensity scales include PREDITOP [102] and PEOPLE [103] (Table 2). PREDITOP [102] uses a multiparametric algorithm based on hydrophilicity, accessibility, flexibility, and secondary structure properties of the amino acids. PEOPLE [103] uses the same parameters and in addition includes the assessment of β-turns. A related method to predict B-cell epitopes was introduced by Kolaskar and Tongaonkar [104], consisting on a simple antigenicity scale derived from physicochemical properties and frequencies of amino acids in experimentally determined B-cell epitopes. This index is perhaps the most popular antigenic scale for B-cell epitope prediction, and it is actually implemented by GCG [105] and EMBOSS [106] packages. Comparative evaluations of propensity scales carried out in a dataset of 85 linear B-cell epitopes showed that most propensity scales predicted between 50 and 70% of B-cell epitopes, with the β-turn scale reaching the best values [101, 107]. It has also been shown that combining the different scales does not appear to improve predictions [102, 108]. Moreover, Blythe and Flower [109] demonstrated that single-scale amino acid propensity scales are not reliable to predict epitope location.

The poor performance of amino acid scales for the prediction of linear B-cell epitopes prompted the introduction of machine learning- (ML-) based methods (Table 2). These methods are developed by training ML algorithms to distinguish experimental B-cell epitopes from non-B-cell epitopes. Prior to training, B-cell epitopes are translated into feature vectors capturing selected properties, such as those given by different propensity scales. Relevant examples of B-cell epitope prediction methods based on ML include BepiPred [110], ABCpred [111], LBtope [112], BCPREDS [113], and SVMtrip [114]. Datasets, training features, and algorithms used for developing these methods differ. BepiPred is based on random forests trained on B-cell epitopes obtained from 3D-structures of antigen-antibody complexes [110]. Both BCPREDS [113] and SVMtrip [114] are based on support vector machines (SVM) but while BCPREDS was trained using various string kernels that eliminate the need for representing the sequence into length-fixed feature vectors, SMVtrip was trained on length-fixed tripeptide composition vectors. ABCpred and LBtope methods consist on artificial neural networks (ANNs) trained on similar positive data, B-cell epitopes, but differ on negative data, non-B-cell epitopes. Negative data used for training ABCpred consisted on random peptides while negative data used for LBtope was based on experimentally validated non-B-cell epitopes form IEDB [15]. In general, B-cell epitope prediction methods employing ML-algorithm are reported to outperform those based on amino acid propensity scales. Nevertheless, some authors have reported that ML algorithms show little improvement over single-scale-based methods [115].

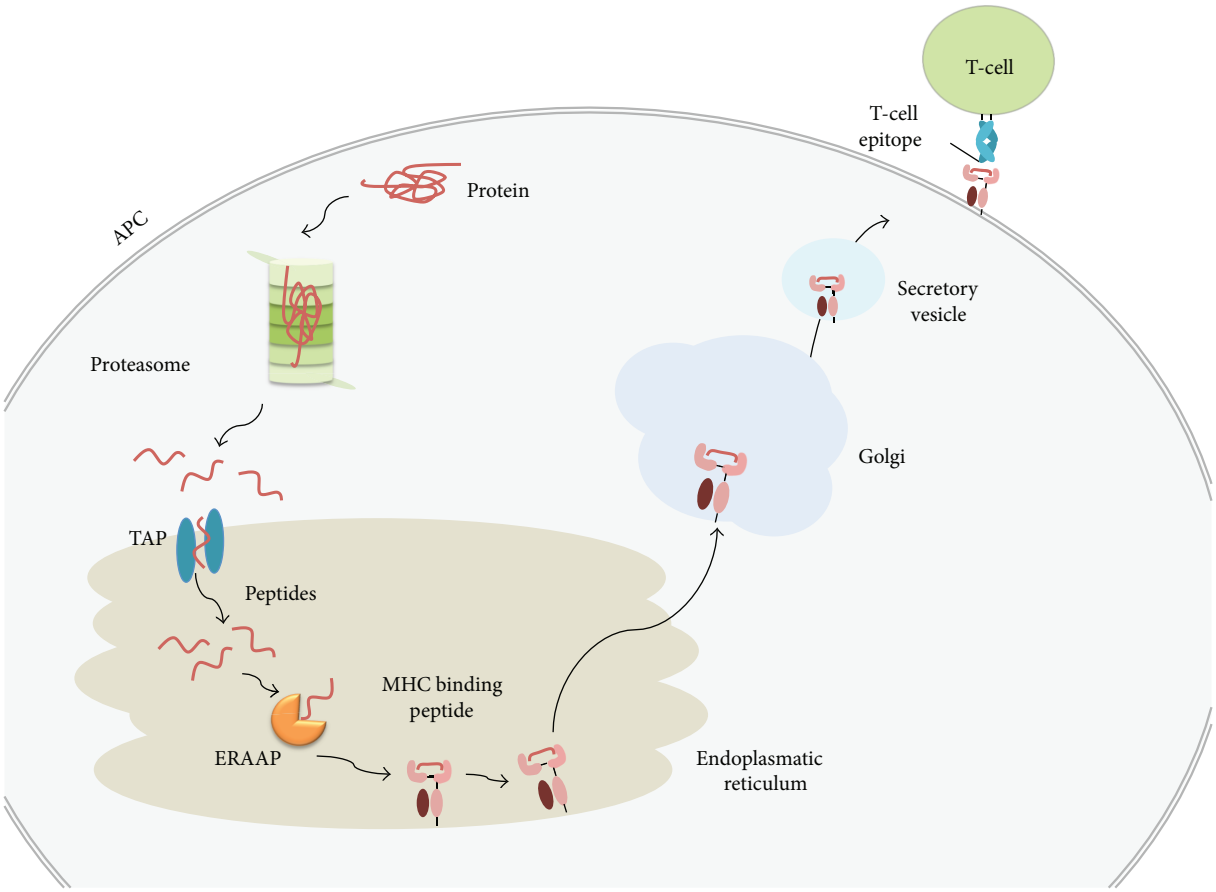


Figure 4: Class I antigen processing. The figure depicts the major steps involved in antigen presentation by MHC I molecules. Proteins are degraded by the proteasome and peptide fragments transported to the endoplasmic reticulum (ER) by TAP where they are loaded onto nascent MHC I molecules. TAP transports peptides ranging from 8 to 16 amino acids. Long peptides cannot bind MHC I molecules but often become suitable for binding after N-terminal trimming by ERAAP.

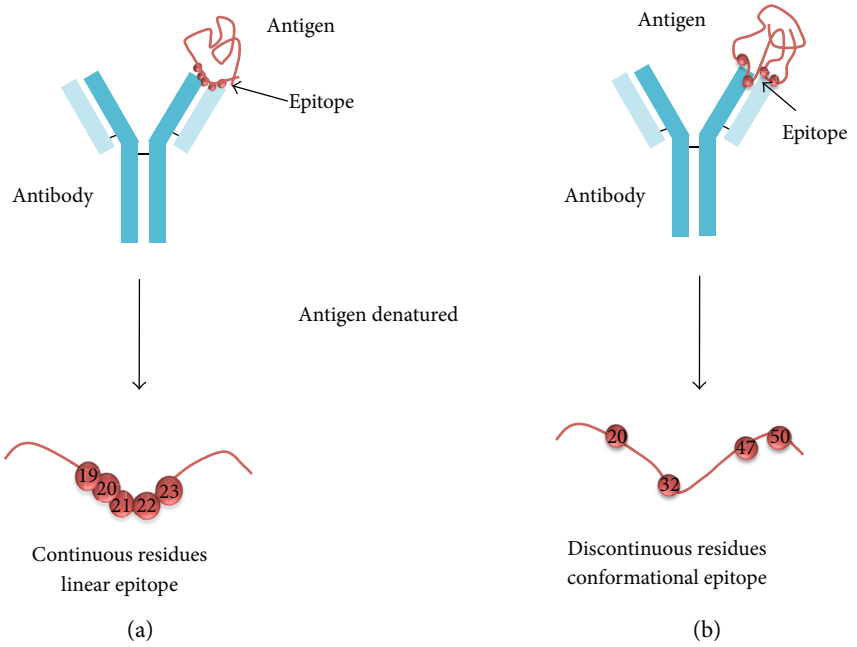


Figure 5: Linear and conformational B-cell epitopes. Linear B-cell epitopes (a) are composed of sequential/continuous residues, while conformational B-cell epitopes (b) contain scattered/discontinuous residues along the sequence.

Table 2: Selected B-cell epitope prediction methods available for free online use.

Tool	Method	Server (URL)	Ref.
Linear B cell epitope			
PEOPLE	Propensity scale method	http://www.iedb.org/	[103]
BepiPred	ML (DT)	http://www.cbs.dtu.dk/services/BepiPred/	[110]
ABCpred	ML (ANN)	http://www.imtech.res.in/raghava/abcpred/	[111]
LBtope	ML (ANN)	http://www.imtech.res.in/raghava/lbtope/	[112]
BCPREDS	ML (SVM)	http://ailab.ist.psu.edu/bcpred/	[113]
SVMtrip	ML (SVM)	http://sysbio.unl.edu/SVMTriP/prediction.php	[114]
Conformational B-cell epitope			
CEP	Structure-based method (solvent accessibility)	http://bioinfo.ernet.in/cep.htm	[118]
DiscoTope	Structure-based method (surface accessibility and propensity amino acid score)	http://tools.iedb.org/discotope/	[119]
ElliPro	Structure-based method (geometrical properties)	http://tools.iedb.org/ellipro/	[121]
PEPITO	Structure-based method (physicochemical properties and geometrical structure)	http://pepito.proteomics.ics.uci.edu/	[122]
SEPPA	Structure-based method (physicochemical properties and geometrical structure)	http://lifecenter.sgst.cn/seppa/	[123]
EPITOPIA	Structure-based method (ML-naïve Bayes)	http://epitopia.tau.ac.il/	[125]
EPSVR	Structure-based method (ML-SVR)	http://sysbio.unl.edu/EPSVR/	[126]
EPIPRED	Structure-based method (ASEP, Docking)	http://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/EpiPred.php	[129]
PEASE	Structure-based method (ASEP, ML)	http://www.ofranlab.org/PEASE	[130]
MIMOX	Mimotope	http://immunet.cn/mimox/helps.html	[131]
PEPITOPE	Mimotope	http://pepitope.tau.ac.il/	[132]
EpiSearch	Mimotope	http://curie.utmb.edu/episearch.html	[133]
MIMOPRO	Mimotope	http://informatics.nenu.edu.cn/MimoPro	[134]
CBTOPE	Sequence based (SVM)	http://www.imtech.res.in/raghava/cbtope/submit.php	[136]

Antibodies elicited in the course of an immune response are generally of a given isotype that determines their biological function. A recent advance in B-cell epitope prediction is the development of a method by Gupta et al. [116] that allows the identification of B-cell epitopes capable of inducing specific class of antibodies. This method is based on SMVs trained on a dataset that includes linear B-cell epitopes known to induce IgG, IgE, and IgA antibodies.

3.2. Prediction of Conformational B-Cell Epitopes. Most B-cell epitopes are conformational and yet, prediction of conformational B-cell epitopes has lagged behind that of linear B-cell epitopes. There are two main practical reasons for that. First of all, prediction of conformational B-cell epitopes generally requires the knowledge of protein three-dimensional (3D) structure and this information is only available for a fraction of proteins [117]. Secondly, isolating conformational B-cell epitopes from their protein context for selective antibody production is a difficult task that requires suitable scaffolds for epitope grafting. Thereby, prediction of conformational B-cell prediction is currently of little relevance for epitope vaccine design and antibody-based technologies. Nonetheless, prediction of conformational B-cell epitopes is interesting for carrying structure-function studies involving antibody-antigen interactions.

There are several available methods to predict conformational B-cell epitopes (Table 2). The first to be introduced was CEP [118], which relied almost entirely on predicting patches of solvent-exposed residues. It was followed by DiscoTope [119], which, in addition to solvent accessibility, considered amino acid statistics and spatial information to predict conformational B-cell epitopes. An independent evaluation of these two methods using a benchmark dataset of 59 conformational epitopes revealed that they did not exceed a 40% of precision and a 46% of recall [120]. Subsequently, more methods were developed, like ElliPro [121] that aims to identify protruding regions in antigen surfaces and PEPITO [122] and SEPPA [123] that combine single physicochemical properties of amino acids and geometrical structure properties. The reported area under the curve (AUC) of these methods is around 0.7, which is indicative of a poor discrimination capacity yet better than random. Though, in an independent evaluation, SEPPA reached an AUC of 0.62 while all the mentioned methods had an AUC around 0.5 [124]. ML has also been applied to predict conformational B-cell epitopes in 3D-structures. Relevant examples include EPITOPIA [125] and EPSVR [126] which are based on naïve Bayes classifiers and support vector regressions, respectively, trained on feature vectors combining different scores. The reported AUC of these two methods is around 0.6.

The above methods for conformational B-cell epitope prediction identify generic antigenic regions regardless of antibodies, which are ignored [127]. However, there are also methods for antibody-specific epitope prediction. This approach was pioneered by Soga et al. [128] who defined an antibody-specific epitope propensity (ASEP) index after analyzing the interfaces of antigen-antibody 3D-structures. Using this index, they developed a novel method for predicting epitope residues in individual antibodies that worked by narrowing down candidate epitope residues predicted by conventional methods. More recently, Krawczyk et al. [129] developed EpiPred, a method that uses a docking-like approach to match up antibody and antigen structures, thus identifying epitope regions on the antigen. A similar approach is used by PEASE [130], adding that this method utilizes the sequence of the antibody and the 3D-structure of the antigen. Briefly, for each pair of antibody sequence and antigen structure, PEASE uses a machine learning model trained on properties from 120 antibody-antigen complexes to identify pair combination of residues from complementarity-determining regions (CDRs) of the antibody and the antigen that are likely to interact.

Another approach to identify conformational B-cell epitopes in a protein with a known 3D-structure is through mimotope-based methods. Mimotopes are peptides selected from randomized peptide libraries for their ability to bind to an antibody raised against a native antigen. Mimotope-based methods require to input antibody affinity-selected peptides and the 3D-structure of the selected antigen. Examples of bioinformatics tools for conformational B-cell epitope prediction using mimotopes include MIMOX [131], PEPITOPE [132], EPISEARCH [133], MIMOPRO [134], and PEPMAPPER [135] (Table 2).

As remarked before, methods for conformational B-cell epitope prediction generally require the 3D-structure of the antigen. Exceptionally, however, Ansari and Raghava [136] developed a method (CBTOPE) for the identification of conformational B-cell epitope from the primary sequence of the antigen. CBTOPE is based on SVM and trained on physicochemical and sequence-derived features of conformational B-cell epitopes. CBTOPE reported accuracy was 86.6% in crossvalidation experiments.

4. Concluding Remarks

Currently, T-cell epitope prediction is more advanced and reliable than that of B-cell prediction. However, while it is possible to confirm experimentally the predicted binding to MHC molecules of most peptides predicted, only ~10% of those are shown to be immunogenic (able to elicit a T-cell response) [68]. Such a low T-cell epitope discovery rate is due to the fact that we do not have adequate models for predicting antigen processing yet [68]. The economic toll of low T-cell epitope discovery rate can be overcome, at least in part, by prioritizing protein antigens for epitope prediction [137–139]. For T-cell epitope vaccine development, researchers can also resort to experimentally known T-cell epitopes, available in epitope databases, selecting through immunoinformatics those that provide maximum population protection coverage [64, 140, 141]. In any case, T-cell epitope prediction remains an integral part of T-cell epitope mapping approaches. In contrast, B-cell epitope prediction utility is currently much more limited. There are several reasons to that. First of all, prediction of B-cell epitopes is still unreliable for both linear and conformational B-cell epitopes. Secondly, linear B-cell epitopes do usually elicit antibodies that do not crossreact with native antigens. Third, the great majority of B-cell epitopes are conformational and yet predicting conformational epitopes have few applications, as they cannot be isolated from their protein context. Under this scenario, the key is not only to improve current methods for B-cell epitope prediction but also to develop novel approaches and platforms for epitope grafting onto suitable scaffolds capable of replacing the native antigen.

To conclude, we wish to make two final remarks that are relevant for epitope vaccine design. First of all, it is that epitope prediction methods can provide potential epitopes from any given protein query but not all the antigens are equally relevant for vaccine development. Therefore, researchers have also developed tools to identify vaccine candidate antigens [142, 143], those likely to induce protective immunity, which can then be targeted for epitope prediction and epitope vaccine design. Second, it should be borne in mind that epitope peptides exhibit little immunogenicity and need to be used in combination with adjuvants, which increase immunogenicity by inducing strong innate immune responses that enable adaptive immunity [144–146]. Consequently, the discovery of new adjuvants is particularly relevant for epitope-based vaccines [146] and to that end, Nagpal et al. [147] developed a pioneered method that can predict the immunomodulatory activity of RNA sequences.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Jose L. Sanchez-Trincado and Marta Gomez-Perosanz contributed equally to this work.

Acknowledgments

The authors wish to thank *Inmunotek*, *SL* and the Spanish Department of Science at MINECO for supporting the Immunomedicine group research through Grants SAF2006: 07879, SAF2009:08301, and BIO2014:54164-R to Pedro A. Reche. The authors also wish to thank Dr. Esther M. Lafuente for critical reading and corrections.

References

[1] W. E. Paul, *Fundamental Immunology*, Lippincott Williams & Wilkins, 2012.

[2] B. Sun and Y. Zhang, "Overview of orchestration of CD4+ T cell subsets in immune responses," *Advances in Experimental Medicine and Biology*, vol. 841, pp. 1–13, 2014.

[3] M. H. Van Regenmortel, "What is a B-cell epitope?," *Methods in Molecular Biology*, vol. 524, pp. 3–20, 2009.

[4] J. Ponomarenko and M. Van Regenmortel, "B-cell epitope prediction," in *Structural Bioinformatics*, pp. 849–879, John Wiley & Sons, Inc, 2009.

[5] T. A. Ahmad, A. E. Eweida, and L. H. El-Sayed, "T-cell epitope mapping for the design of powerful vaccines," *Vaccine Reports*, vol. 6, pp. 13–22, 2016.

[6] L. Malherbe, "T-cell epitope mapping," *Annals of Allergy, Asthma & Immunology*, vol. 103, no. 1, pp. 76–79, 2009.

[7] R. K. Ahmed and M. J. Maeurer, "T-cell epitope mapping," *Methods in Molecular Biology*, vol. 524, pp. 427–438, 2009.

[8] E. M. Lafuente and P. A. Reche, "Prediction of MHC-peptide binding: a systematic and comprehensive overview," *Current Pharmaceutical Design*, vol. 15, no. 28, pp. 3209–3220, 2009.

[9] P. E. Jensen, "Recent advances in antigen processing and presentation," *Nature Immunology*, vol. 8, no. 10, pp. 1041–1048, 2007.

[10] L. J. Stern and D. C. Wiley, "Antigenic peptide binding by class I and class II histocompatibility proteins," *Structure*, vol. 2, no. 4, pp. 245–251, 1994.

[11] D. R. Madden, "The three-dimensional structure of peptide-MHC complexes," *Annual Review of Immunology*, vol. 13, no. 1, pp. 587–622, 1995.

[12] D. R. Madden, D. N. Garboczi, and D. C. Wiley, "The antigenic identity of peptide-MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2," *Cell*, vol. 75, no. 4, pp. 693–708, 1993.

[13] D. V. Desai and U. Kulkarni-Kale, "T-cell epitope prediction methods: an overview," *Methods in Molecular Biology*, vol. 1184, pp. 333–364, 2014.

[14] A. Patronov and I. Doytchinova, "T-cell epitope vaccine design by immunoinformatics," *Open Biology*, vol. 3, no. 1, article 120139, 2013.

[15] R. Vita, J. A. Overton, J. A. Greenbaum et al., "The immune epitope database (IEDB) 3.0," *Nucleic Acids Research*, vol. 43, D1, pp. D405–D412, 2015.

[16] M. Molero-Abraham, E. M. Lafuente, and P. Reche, "Customized predictions of peptide-MHC binding and T-cell epitopes using EPIMHC," *Methods in Molecular Biology*, vol. 1184, pp. 319–332, 2014.

[17] C. P. Toseland, D. J. Clayton, H. McSparron et al., "AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data," *Immunome Research*, vol. 1, no. 1, p. 4, 2005.

[18] S. P. Singh and B. N. Mishra, "Major histocompatibility complex linked databases and prediction tools for designing vaccines," *Human Immunology*, vol. 77, no. 3, pp. 295–306, 2016.

[19] J. D'Amaro, J. G. A. Houbiers, J. W. Drijfhout et al., "A computer program for predicting possible cytotoxic T lymphocyte epitopes based on HLA class I peptide-binding motifs," *Human Immunology*, vol. 43, no. 1, pp. 13–18, 1995.

[20] M. Bouvier and D. Wiley, "Importance of peptide amino and carboxyl termini to the stability of MHC class I molecules," *Science*, vol. 265, no. 5170, pp. 398–402, 1994.

[21] J. Ruppert, J. Sidney, E. Celis, R. T. Kubo, H. M. Grey, and A. Sette, "Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules," *Cell*, vol. 74, no. 5, pp. 929–937, 1993.

[22] M. Nielsen, C. Lundegaard, P. Worning et al., "Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach," *Bioinformatics*, vol. 20, no. 9, pp. 1388–1397, 2004.

[23] H. G. Rammensee, J. Bachmann, N. P. N. Emmerich, O. A. Bachor, and S. Stevanovic, "SYFPEITHI: database for MHC ligands and peptide motifs," *Immunogenetics*, vol. 50, no. 3-4, pp. 213–219, 1999.

[24] P. A. Reche, J. P. Glutting, and E. L. Reinherz, "Prediction of MHC class I binding peptides using profile motifs," *Human Immunology*, vol. 63, no. 9, pp. 701–709, 2002.

[25] P. A. Reche and E. L. Reinherz, "Definition of MHC supertypes through clustering of MHC peptide-binding repertoires," *Methods in Molecular Biology*, vol. 409, pp. 163–173, 2007.

[26] P. A. Reche, J. P. Glutting, H. Zhang, and E. L. Reinherz, "Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles," *Immunogenetics*, vol. 56, no. 6, pp. 405–419, 2004.

[27] M. Gribskov and S. Veretnik, "Identification of sequence pattern with profile analysis," *Methods in Enzymology*, vol. 266, pp. 198–212, 1996.

[28] K. C. Parker, M. A. Bednarek, and J. E. Coligan, "Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains," *The Journal of Immunology*, vol. 152, no. 1, pp. 163–175, 1994.

[29] H. H. Bui, J. Sidney, B. Peters et al., "Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications," *Immunogenetics*, vol. 57, no. 5, pp. 304–314, 2005.

[30] M. Nielsen, C. Lundegaard, and O. Lund, "Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method," *BMC Bioinformatics*, vol. 8, no. 1, p. 238, 2007.

[31] B. Peters and A. Sette, "Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method," *BMC Bioinformatics*, vol. 6, no. 1, p. 132, 2005.

[32] T. Sturniolo, E. Bono, J. Ding et al., "Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices," *Nature Biotechnology*, vol. 17, no. 6, pp. 555–561, 1999.

[33] B. Peters, W. Tong, J. Sidney, A. Sette, and Z. Weng, "Examining the independent binding assumption for binding of peptide epitopes to MHC-I molecules," *Bioinformatics*, vol. 19, no. 14, pp. 1765–1772, 2003.

[34] P. Guan, I. A. Doytchinova, C. Zygouri, and D. R. Flower, "MHCPred: a server for quantitative prediction of peptide–MHC binding," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3621–3624, 2003.

[35] M. Milik, D. Sauer, A. P. Brunmark et al., "Application of an artificial neural network to predict specific class I MHC binding peptide sequences," *Nature Biotechnology*, vol. 16, no. 8, pp. 753–756, 1998.

[36] V. Brusic, G. Rudy, G. Honeyman, J. Hammer, and L. Harrison, "Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network," *Bioinformatics*, vol. 14, no. 2, pp. 121–130, 1998.

[37] P. Donnes and O. Kohlbacher, "Integrated modeling of the major events in the MHC class I antigen processing pathway," *Protein Science*, vol. 14, no. 8, pp. 2132–2140, 2005.

[38] M. Bhasin and G. P. S. Raghava, "SVM based method for predicting HLA-DRB1*0401 binding peptides in an antigen sequence," *Bioinformatics*, vol. 20, no. 3, pp. 421–423, 2004.

[39] L. Jacob and J. P. Vert, "Efficient peptide-MHC-I binding prediction for alleles with few known binders," *Bioinformatics*, vol. 24, no. 3, pp. 358–366, 2008.

[40] S. Zhu, K. Udaka, J. Sidney, A. Sette, K. F. Aoki-Kinoshita, and H. Mamitsuka, "Improving MHC binding peptide prediction by incorporating binding data of auxiliary MHC molecules," *Bioinformatics*, vol. 22, no. 13, pp. 1648–1655, 2006.

[41] C. J. Savoie, N. Kamikawaji, T. Sasazuki, and S. Kuhara, "Use of BONSAI decision trees for the identification of potential MHC class I peptide epitope motifs," *Pacific Symposium on Biocomputing*, vol. 4, pp. 182–189, 1999.

[42] H. Mamitsuka, "Predicting peptides that bind to MHC molecules using supervised learning of hidden Markov models," *Proteins*, vol. 33, no. 4, pp. 460–474, 1998.

[43] C. Zhang, M. G. Bickis, F. X. Wu, and A. J. Kusalik, "Optimally-connected hidden Markov models for predicting MHC-binding peptides," *Journal of Bioinformatics and Computational Biology*, vol. 04, no. 05, pp. 959–980, 2006.

[44] M. Lacerda, K. Scheffler, and C. Seoighe, "Epitope discovery with phylogenetic hidden Markov models," *Molecular Biology and Evolution*, vol. 27, no. 5, pp. 1212–1220, 2010.

[45] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison, *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, 1998.

[46] W. Liu, X. Meng, Q. Xu, D. R. Flower, and T. Li, "Quantitative prediction of mouse class I MHC peptide binding affinity using support vector machine regression (SVR) models," *BMC Bioinformatics*, vol. 7, no. 1, p. 182, 2006.

[47] D. R. Jandrlic, "SVM and SVR-based MHC-binding prediction using a mathematical presentation of peptide sequences," *Computational Biology and Chemistry*, vol. 65, pp. 117–127, 2016.

[48] S. Buus, S. L. Lauemoller, P. Worning et al., "Sensitive quantitative predictions of peptide-MHC binding by a 'query by committee' artificial neural network approach," *Tissue Antigens*, vol. 62, no. 5, pp. 378–384, 2003.

[49] M. Nielsen, C. Lundegaard, P. Worning et al., "Reliable prediction of T-cell epitopes using neural networks with novel sequence representations," *Protein Science*, vol. 12, no. 5, pp. 1007–1017, 2003.

[50] M. Nielsen and O. Lund, "NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction," *BMC Bioinformatics*, vol. 10, no. 1, p. 296, 2009.

[51] K. Yu, N. Petrovsky, C. Schonbach, J. Y. Koh, and V. Brusic, "Methods for prediction of peptide binding to MHC molecules: a comparative study," *Molecular Medicine*, vol. 8, no. 3, pp. 137–148, 2002.

[52] P. Wang, J. Sidney, C. Dow, B. Mothe, A. Sette, and B. Peters, "A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach," *PLoS Computational Biology*, vol. 4, no. 4, article e1000048, 2008.

[53] P. A. Reche and E. L. Reinherz, "Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms," *Journal of Molecular Biology*, vol. 331, no. 3, pp. 623–641, 2003.

[54] M. Nielsen, C. Lundegaard, T. Blicher et al., "*NetMHCpan*, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence," *PLoS One*, vol. 2, no. 8, article e796, 2007.

[55] M. Nielsen, C. Lundegaard, T. Blicher et al., "Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan," *PLoS Computational Biology*, vol. 4, no. 7, article e1000107, 2008.

[56] P. I. Terasaki, "A brief history of HLA," *Immunologic Research*, vol. 38, no. 1-3, pp. 139–148, 2007.

[57] A. Sette and J. Sidney, "HLA supertypes and supermotifs: a functional perspective on HLA polymorphism," *Current Opinion in Immunology*, vol. 10, no. 4, pp. 478–482, 1998.

[58] A. Sette and J. Sidney, "Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism," *Immunogenetics*, vol. 50, no. 3-4, pp. 201–212, 1999.

[59] I. A. Doytchinova and D. R. Flower, "In silico identification of supertypes for class II MHCs," *The Journal of Immunology*, vol. 174, no. 11, pp. 7085–7095, 2005.

[60] J. Greenbaum, J. Sidney, J. Chung, C. Brander, B. Peters, and A. Sette, "Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes," *Immunogenetics*, vol. 63, no. 6, pp. 325–335, 2011.

[61] O. Lund, M. Nielsen, C. Kesmir et al., "Definition of supertypes for HLA molecules using clustering of specificity matrices," *Immunogenetics*, vol. 55, no. 12, pp. 797–810, 2004.

[62] G. L. Zhang, D. S. DeLuca, D. B. Keskin et al., "MULTI-PRED2: a computational system for large-scale identification of peptides predicted to bind to HLA supertypes and alleles," *Journal of Immunological Methods*, vol. 374, no. 1-2, pp. 53–61, 2011.

[63] P. A. Reche and E. L. Reinherz, "PEPVAC: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands," *Nucleic Acids Research*, vol. 33, Supplement 2, pp. W138–W142, 2005.

[64] M. Molero-Abraham, E. M. Lafuente, D. R. Flower, and P. A. Reche, "Selection of conserved epitopes from hepatitis C virus for pan-populational stimulation of T-cell responses," *Clinical and Developmental Immunology*, vol. 2013, Article ID 601943, 10 pages, 2013.

[65] S. L. Constant and K. Bottomly, "Induction of Th1 and Th2 $CD4^+$ T cell responses: the alternative approaches," *Annual Review of Immunology*, vol. 15, no. 1, pp. 297–322, 1997.

[66] E. M. Janssen, A. J. M. van Oosterhout, A. J. M. L. van Rensen, W. van Eden, F. P. Nijkamp, and M. H. M. Wauben, "Modulation of Th2 responses by peptide analogues in a murine model of allergic asthma: amelioration or deterioration of the disease process depends on the Th1 or Th2 skewing characteristics of the therapeutic peptide," *The Journal of Immunology*, vol. 164, no. 2, pp. 580–588, 2000.

[67] S. K. Dhanda, S. Gupta, P. Vir, and G. P. Raghava, "Prediction of IL4 inducing peptides," *Clinical and Developmental Immunology*, vol. 2013, Article ID 263952, 9 pages, 2013.

[68] W. Zhong, P. A. Reche, C. C. Lai, B. Reinhold, and E. L. Reinherz, "Genome-wide characterization of a viral cytotoxic T lymphocyte epitope repertoire," *The Journal of Biological Chemistry*, vol. 278, no. 46, pp. 45135–45144, 2003.

[69] J. S. Blum, P. A. Wearsch, and P. Cresswell, "Pathways of antigen processing," *Annual Review of Immunology*, vol. 31, no. 1, pp. 443–473, 2013.

[70] E. Hoze, L. Tsaban, Y. Maman, and Y. Louzoun, "Predictor for the effect of amino acid composition on CD4 + T cell

epitopes preprocessing," *Journal of Immunological Methods*, vol. 391, no. 1-2, pp. 163–173, 2013.

[71] G. E. Hammer, F. Gonzalez, M. Champsaur, D. Cado, and N. Shastri, "The aminopeptidase ERAAP shapes the peptide repertoire displayed by major histocompatibility complex class I molecules," *Nature Immunology*, vol. 7, no. 1, pp. 103–112, 2006.

[72] A. K. Nussbaum, C. Kuttler, K. P. Hadeler, H. G. Rammensee, and H. Schild, "PAProC: a prediction algorithm for proteasomal cleavages available on the WWW," *Immunogenetics*, vol. 53, no. 2, pp. 87–94, 2001.

[73] H. G. Holzhutter, C. Frommel, and P. M. Kloetzel, "A theoretical approach towards the identification of cleavage-determining amino acid motifs of the 20s proteasome," *Journal of Molecular Biology*, vol. 286, no. 4, pp. 1251–1265, 1999.

[74] M. Nielsen, C. Lundegaard, O. Lund, and C. Kesmir, "The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage," *Immunogenetics*, vol. 57, no. 1-2, pp. 33–41, 2005.

[75] C. M. Diez-Rivero, E. M. Lafuente, and P. A. Reche, "Computational analysis and modeling of cleavage by the immunoproteasome and the constitutive proteasome," *BMC Bioinformatics*, vol. 11, no. 1, p. 479, 2010.

[76] M. Bhasin and G. P. S. Raghava, "Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences," *Nucleic Acids Research*, vol. 33, Supplement 1, pp. W202–W207, 2005.

[77] M. Bhasin and G. P. Raghava, "Analysis and prediction of affinity of TAP binding peptides using cascade SVM," *Protein Science*, vol. 13, no. 3, pp. 596–607, 2004.

[78] C. M. Diez-Rivero, B. Chenlo, P. Zuluaga, and P. A. Reche, "Quantitative modeling of peptide binding to TAP using support vector machine," *Proteins*, vol. 78, no. 1, pp. 63–72, 2010.

[79] S. Daniel, V. Brusic, S. Caillat-Zucman et al., "Relationship between peptide selectivities of human transporters associated with antigen processing and HLA class I molecules," *The Journal of Immunology*, vol. 161, no. 2, pp. 617–624, 1998.

[80] V. Brusic, P. van Endert, J. Zeleznikow, S. Daniel, J. Hammer, and N. Petrovsky, "A neural network model approach to the study of human TAP transporter," *In Silico Biology*, vol. 1, no. 2, pp. 109–121, 1999.

[81] S. Tenzer, B. Peters, S. Bulik et al., "Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding," *Cellular and Molecular Life Sciences*, vol. 62, no. 9, pp. 1025–1037, 2005.

[82] I. A. Doytchinova, P. Guan, and D. R. Flower, "EpiJen: a server for multistep T cell epitope prediction," *BMC Bioinformatics*, vol. 7, no. 1, p. 131, 2006.

[83] M. V. Larsen, C. Lundegaard, K. Lamberth et al., "An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions," *European Journal of Immunology*, vol. 35, no. 8, pp. 2295–2303, 2005.

[84] B. Schubert, H. P. Brachvogel, C. Jurges, and O. Kohlbacher, "EpiToolKit—a web-based workbench for vaccine design," *Bioinformatics*, vol. 31, no. 13, pp. 2211–2213, 2015.

[85] B. Schubert, M. Walzer, H. P. Brachvogel, A. Szolek, C. Mohr, and O. Kohlbacher, "FRED 2: an immunoinformatics framework for Python," *Bioinformatics*, vol. 32, no. 13, pp. 2044–2046, 2016.

[86] M. Atanasova, A. Patronov, I. Dimitrov, D. R. Flower, and I. Doytchinova, "EpiDOCK: a molecular docking-based tool for MHC class II binding prediction," *Protein Engineering, Design and Selection*, vol. 26, no. 10, pp. 631–634, 2013.

[87] J. Hakenberg, A. K. Nussbaum, H. Schild et al., "MAPPP: MHC class I antigenic peptide processing prediction," *Applied Bioinformatics*, vol. 2, no. 3, pp. 155–158, 2003.

[88] P. Oyarzun, J. J. Ellis, M. Boden, and B. Kobe, "PREDIVAC: CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II DR protein diversity," *BMC Bioinformatics*, vol. 14, no. 1, p. 52, 2013.

[89] Y. He, Z. Xiang, and H. L. T. Mobley, "Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development," *Journal of Biomedicine and Biotechnology*, vol. 2010, Article ID 297505, 15 pages, 2010.

[90] I. Dimitrov, P. Garnev, D. R. Flower, and I. Doytchinova, "EpiTOP—a proteochemometric tool for MHC class II binding prediction," *Bioinformatics*, vol. 26, no. 16, pp. 2066–2068, 2010.

[91] H. Singh and G. P. S. Raghava, "ProPred: prediction of HLA-DR binding sites," *Bioinformatics*, vol. 17, no. 12, pp. 1236-1237, 2001.

[92] H. Singh and G. P. Raghava, "ProPred1: prediction of promiscuous MHC class-I binding sites," *Bioinformatics*, vol. 19, no. 8, pp. 1009–1014, 2003.

[93] Q. Zhang, P. Wang, Y. Kim et al., "Immune epitope database analysis resource (IEDB-AR)," *Nucleic Acids Research*, vol. 36, Web Server issue, pp. W513–W518, 2008.

[94] M. Bhasin and G. P. S. Raghava, "A hybrid approach for predicting promiscuous MHC class I restricted T cell epitopes," *Journal of Biosciences*, vol. 32, no. 1, pp. 31–42, 2007.

[95] P. Donnes and A. Elofsson, "Prediction of MHC class I binding peptides, using SVMHC," *BMC Bioinformatics*, vol. 3, no. 1, p. 25, 2002.

[96] T. P. Hopp and K. R. Woods, "Prediction of protein antigenic determinants from amino acid sequences," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 78, no. 6, pp. 3824–3828, 1981.

[97] T. P. Hopp and K. R. Woods, "A computer program for predicting protein antigenic determinants," *Molecular Immunology*, vol. 20, no. 4, pp. 483–489, 1983.

[98] L. Lins, A. Thomas, and R. Brasseur, "Analysis of accessible surface of residues in proteins," *Protein Science*, vol. 12, no. 7, pp. 1406–1417, 2003.

[99] P. A. Karplus and G. E. Schulz, "Prediction of chain flexibility in proteins: a tool for the selection of peptide antigen," *Naturwissenschaften*, vol. 72, no. 4, pp. 212-213, 1985.

[100] E. A. Emini, J. V. Hughes, D. S. Perlow, and J. Boger, "Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide," *Journal of Virology*, vol. 55, no. 3, pp. 836–839, 1985.

[101] J. L. Pellequer, E. Westhof, and M. H. V. Van Regenmortel, "Correlation between the location of antigenic sites and the prediction of turns in proteins," *Immunology Letters*, vol. 36, no. 1, pp. 83–99, 1993.

[102] J. L. Pellequer and E. Westhof, "PREDITOP: a program for antigenicity prediction," *Journal of Molecular Graphics*, vol. 11, no. 3, pp. 204–210, 1993, 191–2.

[103] A. J. P. Alix, "Predictive estimation of protein linear epitopes by using the program PEOPLE," *Vaccine*, vol. 18, no. 3-4, pp. 311–314, 1999.

[104] A. S. Kolaskar and P. C. Tongaonkar, "A semi-empirical method for prediction of antigenic determinants on protein antigens," *FEBS Letters*, vol. 276, no. 1-2, pp. 172–174, 1990.

[105] D. D. Womble, "GCG: the Wisconsin package of sequence analysis programs," *Methods in Molecular Biology*, vol. 132, pp. 3–22, 2000.

[106] P. Rice, I. Longden, and A. Bleasby, "EMBOSS: the European molecular biology open software suite," *Trends in Genetics*, vol. 16, no. 6, pp. 276-277, 2000.

[107] J. L. Pellequer, E. Westhof, and M. H. V. Van Regenmortel, "Predicting location of continuous epitopes in proteins from their primary structures," *Methods in Enzymology*, vol. 203, pp. 176–201, 1991.

[108] M. Odorico and J. L. Pellequer, "BEPITOPE: predicting the location of continuous epitopes and patterns in proteins," *Journal of Molecular Recognition*, vol. 16, no. 1, pp. 20–22, 2003.

[109] M. J. Blythe and D. R. Flower, "Benchmarking B cell epitope prediction: underperformance of existing methods," *Protein Science*, vol. 14, no. 1, pp. 246–248, 2005.

[110] M. C. Jespersen, B. Peters, M. Nielsen, and P. Marcatili, "BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes," *Nucleic Acids Research*, vol. 45, Web Server issue, pp. W24–W29, 2017.

[111] S. Saha and G. P. S. Raghava, "Prediction of continuous B-cell epitopes in an antigen using recurrent neural network," *Proteins*, vol. 65, no. 1, pp. 40–48, 2006.

[112] H. Singh, H. R. Ansari, and G. P. S. Raghava, "Improved method for linear B-cell epitope prediction using antigen's primary sequence," *PLoS One*, vol. 8, no. 5, article e62216, 2013.

[113] Y. El-Manzalawy, D. Dobbs, and V. Honavar, "Predicting linear B-cell epitopes using string kernels," *Journal of Molecular Recognition*, vol. 21, no. 4, pp. 243–255, 2008.

[114] B. Yao, L. Zhang, S. Liang, and C. Zhang, "SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity," *PLoS One*, vol. 7, no. 9, article e45152, 2012.

[115] J. A. Greenbaum, P. H. Andersen, M. Blythe et al., "Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools," *Journal of Molecular Recognition*, vol. 20, no. 2, pp. 75–82, 2007.

[116] S. Gupta, H. R. Ansari, A. Gautam, Open Source Drug Discovery Consortium, and G. P. Raghava, "Identification of B-cell epitopes in an antigen for inducing specific class of antibodies," *Biology Direct*, vol. 8, no. 1, p. 27, 2013.

[117] M. Levitt, "Nature of the protein universe," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 27, pp. 11079–11084, 2009.

[118] U. Kulkarni-Kale, S. Bhosle, and A. S. Kolaskar, "CEP: a conformational epitope prediction server," *Nucleic Acids Research*, vol. 33, Web Server issue, pp. W168–W171, 2005.

[119] P. Haste Andersen, M. Nielsen, and O. Lund, "Prediction of residues in discontinuous B-cell epitopes using protein 3D structures," *Protein Science*, vol. 15, no. 11, pp. 2558–2567, 2006.

[120] J. V. Ponomarenko and P. E. Bourne, "Antibody-protein interactions: benchmark datasets and prediction tools evaluation," *BMC Structural Biology*, vol. 7, no. 1, p. 64, 2007.

[121] J. Ponomarenko, H. H. Bui, W. Li et al., "ElliPro: a new structure-based tool for the prediction of antibody epitopes," *BMC Bioinformatics*, vol. 9, no. 1, p. 514, 2008.

[122] M. J. Sweredoski and P. Baldi, "PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure," *Bioinformatics*, vol. 24, no. 12, pp. 1459-1460, 2008.

[123] J. Sun, D. Wu, T. Xu et al., "SEPPA: a computational server for spatial epitope prediction of protein antigens," *Nucleic Acids Research*, vol. 37, no. suppl_2, Web Server issue, pp. W612–W616, 2009.

[124] X. Xu, J. Sun, Q. Liu et al., "Evaluation of spatial epitope computational tools based on experimentally-confirmed dataset for protein antigens," *Chinese Science Bulletin*, vol. 55, no. 20, pp. 2169–2174, 2010.

[125] N. D. Rubinstein, I. Mayrose, E. Martz, and T. Pupko, "Epitopia: a web-server for predicting B-cell epitopes," *BMC Bioinformatics*, vol. 10, no. 1, p. 287, 2009.

[126] S. Liang, D. Zheng, D. M. Standley, B. Yao, M. Zacharias, and C. Zhang, "EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results," *BMC Bioinformatics*, vol. 11, no. 1, p. 381, 2010.

[127] I. Sela-Culang, Y. Ofran, and B. Peters, "Antibody specific epitope prediction — emergence of a new paradigm," *Current Opinion in Virology*, vol. 11, pp. 98–102, 2015.

[128] S. Soga, D. Kuroda, H. Shirai, M. Kobori, and N. Hirayama, "Use of amino acid composition to predict epitope residues of individual antibodies," *Protein Engineering, Design and Selection*, vol. 23, no. 6, pp. 441–448, 2010.

[129] K. Krawczyk, X. Liu, T. Baker, J. Shi, and C. M. Deane, "Improving B-cell epitope prediction and its application to global antibody-antigen docking," *Bioinformatics*, vol. 30, no. 16, pp. 2288–2294, 2014.

[130] I. Sela-Culang, S. Ashkenazi, B. Peters, and Y. Ofran, "PEASE: predicting B-cell epitopes utilizing antibody sequence," *Bioinformatics*, vol. 31, no. 8, pp. 1313–1315, 2015.

[131] J. Huang, A. Gutteridge, W. Honda, and M. Kanehisa, "MIMOX: a web tool for phage display based epitope mapping," *BMC Bioinformatics*, vol. 7, no. 1, p. 451, 2006.

[132] I. Mayrose, O. Penn, E. Erez et al., "Pepitope: epitope mapping from affinity-selected peptides," *Bioinformatics*, vol. 23, no. 23, pp. 3244–3246, 2007.

[133] S. S. Negi and W. Braun, "Automated detection of conformational epitopes using phage display peptide sequences," *Bioinformatics and Biology Insights*, vol. 3, pp. 71–81, 2009.

[134] W. Chen, P. Sun, Y. Lu, W. W. Guo, Y. Huang, and Z. Ma, "MimoPro: a more efficient web-based tool for epitope prediction using phage display libraries," *BMC Bioinformatics*, vol. 12, no. 1, p. 199, 2011.

[135] W. Chen, W. W. Guo, Y. Huang, and Z. Ma, "PepMapper: a collaborative web tool for mapping epitopes from affinity-selected peptides," *PLoS One*, vol. 7, no. 5, article e37869, 2012.

[136] H. Ansari and G. P. S. Raghava, "Identification of conformational B-cell epitopes in an antigen from its primary sequence," *Immunome Research*, vol. 6, no. 1, p. 6, 2010.

[137] C. M. Diez-Rivero and P. A. Reche, "CD8 T cell epitope distribution in viruses reveals patterns of protein biosynthesis," *PLoS One*, vol. 7, no. 8, article e43674, 2012.

[138] D. R. Flower, I. K. Macdonald, K. Ramakrishnan, M. N. Davies, and I. A. Doytchinova, "Computer aided selection of candidate vaccine antigens," *Immunome Research*, vol. 6, Supplement 2, p. S1, 2010.

[139] M. Molero-Abraham, J. P. Glutting, D. R. Flower, E. M. Lafuente, and P. A. Reche, "EPIPOX: immunoinformatic characterization of the shared T-cell epitome between variola virus and related pathogenic orthopoxviruses," *Journal of Immunology Research*, vol. 2015, Article ID 738020, 11 pages, 2015.

[140] P. A. Reche, D. B. Keskin, R. E. Hussey, P. Ancuta, D. Gabuzda, and E. L. Reinherz, "Elicitation from virus-naive individuals of cytotoxic T lymphocytes directed against conserved HIV-1 epitopes," *Medical Immunology*, vol. 5, no. 1, p. 1, 2006.

[141] Q. M. Sheikh, D. Gatherer, P. A. Reche, and D. R. Flower, "Towards the knowledge-based design of universal influenza epitope ensemble vaccines," *Bioinformatics*, vol. 32, no. 21, pp. 3233–3239, 2016.

[142] S. J. Goodswen, P. J. Kennedy, and J. T. Ellis, "*Vacceed*: a high-throughput *in silico* vaccine candidate discovery pipeline for eukaryotic pathogens based on reverse vaccinology," *Bioinformatics*, vol. 30, no. 16, pp. 2381–2383, 2014.

[143] M. Rizwan, A. Naz, J. Ahmad et al., "VacSol: a high throughput in silico pipeline to predict potential therapeutic targets in prokaryotic pathogens using subtractive reverse vaccinology," *BMC Bioinformatics*, vol. 18, no. 1, p. 106, 2017.

[144] H. Yang and D. S. Kim, "Peptide immunotherapy in vaccine development: from epitope to adjuvant," *Advances in Protein Chemistry and Structural Biology*, vol. 99, pp. 1–14, 2015.

[145] A. Di Pasquale, S. Preiss, F. Tavares Da Silva, and N. Garcon, "Vaccine adjuvants: from 1920 to 2015 and beyond," *Vaccines*, vol. 3, no. 2, pp. 320–343, 2015.

[146] F. Azmi, A. A. Ahmad Fuaad, M. Skwarczynski, and I. Toth, "Recent progress in adjuvant discovery for peptide-based subunit vaccines," *Human Vaccines & Immunotherapeutics*, vol. 10, no. 3, pp. 778–796, 2014.

[147] G. Nagpal, K. Chaudhary, S. K. Dhanda, and G. P. S. Raghava, "Computational prediction of the immunomodulatory potential of RNA sequences," *Methods in Molecular Biology*, vol. 1632, pp. 75–90, 2017.

Hindawi
Journal of Immunology Research
Volume 2017, Article ID 9363750, 15 pages
https://doi.org/10.1155/2017/9363750

Research Article

Computer-Aided Design of an Epitope-Based Vaccine against Epstein-Barr Virus

Julio Alonso-Padilla,[1] Esther M. Lafuente,[2] and Pedro A. Reche[2]

[1]*Barcelona Institute for Global Health (ISGlobal), Centre for Research in International Health (CRESIB), Hospital Clinic-University of Barcelona, Barcelona, Spain*
[2]*Laboratory of Immunomedicine, Faculty of Medicine, University Complutense of Madrid, Ave Complutense S/N, 28040 Madrid, Spain*

Correspondence should be addressed to Pedro A. Reche; parecheg@med.ucm.es

Received 19 May 2017; Revised 7 August 2017; Accepted 20 August 2017; Published 28 September 2017

Academic Editor: Peirong Jiao

Epstein-Barr virus is a very common human virus that infects 90% of human adults. EBV replicates in epithelial and B cells and causes infectious mononucleosis. EBV infection is also linked to various cancers, including Burkitt's lymphoma and nasopharyngeal carcinomas, and autoimmune diseases such as multiple sclerosis. Currently, there are no effective drugs or vaccines to treat or prevent EBV infection. Herein, we applied a computer-aided strategy to design a prophylactic epitope vaccine ensemble from experimentally defined T and B cell epitopes. Such strategy relies on identifying conserved epitopes in conjunction with predictions of HLA presentation for T cell epitope selection and calculations of accessibility and flexibility for B cell epitope selection. The T cell component includes 14 CD8 T cell epitopes from early antigens and 4 CD4 T cell epitopes, targeted during the course of a natural infection and providing a population protection coverage of over 95% and 81.8%, respectively. The B cell component consists of 3 experimentally defined B cell epitopes from gp350 plus 4 predicted B cell epitopes from other EBV envelope glycoproteins, all mapping in flexible and solvent accessible regions. We discuss the rationale for the formulation and possible deployment of this epitope vaccine ensemble.

1. Introduction

Epstein-Barr virus (EBV), or human herpesvirus 4, is a large enveloped virus that belongs to the family herpesviruses γ. It has a size of 120–180 nm and a double-stranded linear DNA genome (~171 Kb long), encoding ~90 genes [1]. The genome is enclosed within a nucleocapsid protein which is in turn surrounded by a lipid envelope that contains the viral surface proteins essential for infection [2]. According to its expression, EBV genes are divided into immediate early (expressed very early during lytic infection, coding for transcription factors), early (interfere with the host metabolism and DNA synthesis), and late genes (including structural and nonstructural glycoproteins). There are two major subtypes of EBV (type 1 and type 2), which mainly differ in their nuclear antigen-3 gene (EBNA-3). Both types are detected all over the world, yet type 1 is dominant in most populations [3].

EBV is present in over 90% of the adult world population [4]. Most people become infected with EBV during childhood and develop little or no symptoms. However, if the infection occurs later in life, it can cause infectious mononucleosis (IM) in about 30–50% of the cases [5]. Viral transmission is primarily through saliva; hence, the nickname of kiss disease for IM. The virus can infect and replicate in epithelial and B cells. Infection of epithelial cells of the oropharynx has a relevant role in EBV expansion during primary infection [2]. However, B cells are the main targets of the virus. They are fundamental to establish an EBV infection—X-linked agammaglobulinemic patients are not infected by the virus [6]—and can pass the virus to epithelial cells by direct contact [7]. Moreover, it is in memory B cells that the virus persists as a long-term latent infection [8]. Tropism of EBV for B lymphocytes is mediated by cell surface molecules

CD21 (i.e., complement receptor 2 (CR2)) and HLA-II that serve as receptors of the viral envelope glycoproteins gp350 and gp42, respectively [9]. Infection of B cells by EBV does not usually release viral progeny. Instead, the virus activates the cell cycle driving the expansion of latently infected B cells, inducing its own proliferation, thus getting persistently established in the lymphoid system [7, 8]. Latency is not permanent though, as EBV can periodically switch between latent and lytic states. Reactivation from latency is triggered by environmental stimuli and the process is tightly controlled by the immune system [10].

Immunity against EBV has been studied extensively [10, 11]. Natural killer (NK) cells play an important role in the innate immune response, delaying or preventing the EBV transformation of B cells through the production of interferon gamma (IFN-γ) [12]. Subsequently, the virus elicits strong adaptive immune responses, primarily mediated by cytotoxic CD8 T cells. CD8 T cell responses eliminate viral-infected cells upon recognition of EBV peptide antigens bound to MHC I molecules in the surface of target cells. Cytotoxic CD8 T cell response against EBV infection is so dramatic that, in IM patients, up to 50% of CD8 T cells recognize EBV-specific CD8 T cell epitopes, most derived from immediate early or early antigens [13, 14]. In contrast, CD4 T cell responses against the virus are less dramatic and focused [13]. CD4 T cells recognize peptide antigens bound to MHC II molecules and commit into different phenotypes of cytokine-producing T helper cells (Th) that control the immune response. Most EBV-specific CD4 T cells produce IFN-γ and tumor necrosis factor alpha (TNF-α), with a smaller number producing IL-2 which is the usual and expected Th1 antiviral response [15]. Regarding the humoral immune response, EBV infection triggers a potent reaction against various viral antigens. The acute primary infection is associated with the induction of IgM antibodies against the virus capsid antigen (VCA), which switches to an IgG isotype. IgG anti-VCA antibodies are not neutralizing and remain for life. Neutralizing IgG antibodies targeting viral major glycoprotein gp350 arise only after the resolution of the primary infection [16]. Other antibodies targeting non-neutralizing antigens (e.g., viral proteins located intracellularly) also appear sometime after the resolution of the primary infection [16, 17].

The immune system is capable of controlling EBV primary infection and reactivation phases, forcing the virus to stay latent in memory B cells. Such a control likely has a toll in the immune system. In fact, after extended periods of latency and being facilitated by its potent growth transforming capability, EBV appears to promote an increasing number of human cancers. Frequent cancers linked to EBV include several B cell malignancies, such as Burkitt's lymphoma (BL) and Hodgkin's lymphoma (HL), and epithelial cell malignancies, notably nasopharyngeal carcinoma (NPC) [18]. Furthermore, EBV infection has been implicated with autoimmunity and it is clearly a risk factor for developing multiple sclerosis and to a lesser systemic lupus erythematosus [19].

Currently, no medicine can cure EBV infection and there is no prophylactic or therapeutic vaccine against it. Clearly, a prophylactic vaccine against EBV will have a major impact in public health as it will prevent both EBV infection and related diseases [20]. In this study, we explored a reverse-vaccinology approach to design a prophylactic vaccine against EBV based on CD8 and CD4 T cell epitopes and B cell epitopes. For designing the T cell epitope vaccine component, we relied on combining legacy experimentation with bioinformatics analysis aimed to identify conserved and highly promiscuous T cell epitopes [21–23]. Given the size and complexity of EBV, we also introduced expression criteria to reduce the number of T cell epitopes and focus on those from early antigens with acknowledged function at the initial steps of primary infection [23, 24]. As for the B cell component, we included highly conserved experimentally determined B cell epitopes from EBV gp350 protein as well as potential B cell epitopes predicted in flexible solvent-exposed regions of other envelope proteins important for infection like gp42, gB, and gL. We are confident that our epitope vaccine ensemble poses a basis for developing a powerful and effective vaccine against EBV. Moreover, we trust that the approach and methods introduced in this work ought to become a paradigm of general use in reverse vaccinology.

2. Materials and Methods

2.1. Collection of EBV-Specific Epitopes. We retrieved experimentally defined EBV-specific T and B epitope sequences from the EPIMHC [25] and IEDB [26]. As inclusion criteria, we considered positive assays (excluding low-positive responses) and epitopes being linked to the course of a natural infection in humans for T cell epitopes and any human disease for B cell epitopes. We discarded duplicate peptides and when available, we also retrieved the MHC restriction elements of T cell epitopes. For B cell epitopes, we considered all unique sequences that were not included as part of longer peptides. In total, we obtained 247 unique B cell epitopes and 109 unique T cell epitopes (88 CD8 T cell epitopes and 21 CD4 T cell epitopes). These epitopes are available as supplementary data in Additional File S1 available online at https://doi.org/10.1155/2017/9363750, including Tables S1A, S1B, and S1C for CD8, CD4, and B cell epitopes, respectively. Perl scripts used to identify unique B and T cell epitopes from IEDB search outputs can be obtained from the corresponding author.

2.2. Generation of Clusters and Multiple Sequence Alignments of EBV Protein Sequences. We used CD-HIT [27] with default settings to generate clusters from 13,899 EBV protein sequences that included 89 translated coding DNA sequences (CDS) from a reference genome virus (accession: NC_007605). The protein sequences were downloaded following the links in the NCBI taxonomy database (TAX ID: 10376) [28]. We processed CD-HIT clusters with reference EBV proteins, removed identical sequences, and subsequently generated multiple sequence alignments (MSA) using MUSCLE [29]. As a result, we obtained 85 referenced MSA of EBV proteins that were used for further analysis. Software for clustering the sequences will be provided by the corresponding author upon written request.

2.3. Generation of EBV-Reference Proteome with Variable Sites Masked and Identification of Conserved Epitopes. We generated EBV-reference sequences with variable sites masked upon sequence variability analyses on the referenced MSA of EBV proteins. Briefly, we calculated the sequence variability in the MSA of EBV proteins using the Shannon entropy [30], H, as a variability metric [21, 24, 31]. Shannon entropy per site in a MSA is given by

$$H = -\sum_{i=1}^{M} P_i \log_2 P_i, \quad (1)$$

where Pi is the fraction of residues of amino acid type i and M is equal to 20, the number of amino acid types. H ranges from 0 (total conservation, only one amino acid type is present at that position) to 4.322 (all 20 amino acids are equally represented in that position). We considered gaps as no data. To generate reference EBV consensus sequences, we assigned the computed variability, H, to the EBV-reference proteins included in the MSA and subsequently masked all positions with a variability, H, greater than 0.5 [32, 33]. We used this reference sequence to discard epitope sequences that did not match entirely with it. Hence, the epitopes that we considered conserved did not have a single residue with $H > 0.5$.

2.4. Prediction of Peptide HLA Presentation Profiles and Computation of Population Protection Coverage. T cells only recognize peptides when presented in the cell surface of antigen-presenting cells bound to HLA molecules (MHC molecules in humans). Therefore, we anticipated HLA presentation profiles of peptides by predicting peptide-HLA binding. For CD8 T cell epitopes, we predicted peptide binding using 55 HLA I-specific motif profiles [34–36]. A top 2% rank percentile was used to consider binding to the relevant HLA I molecule. For CD4 T cell epitopes, we predicted peptide binding to 15 reference HLA-DR molecules [37] using the IEDB binding tool [38]. We used a 5% percentile rank cutoff to consider that binding had occurred. The population protection coverage (PPC) of a set of epitopes is the proportion of the population that could elicit an immune response against any of them and can be computed by knowing the gene frequencies of the HLA I alleles that can present the epitopes [21]. For HLA I-restricted T cell epitopes, we used EPISOPT to compute epitope PPC [39]. EPISOPT uses HLA I allele frequencies for 5 distinct ethnic groups in the USA population (Caucasian, Hispanic, Black, Asian, and North American natives) [40] and can identify combinations of epitopes reaching a determined PPC in each of the population groups. We aimed to identify epitope combinations reaching a PPC of 95% in the 5 ethnic groups. For HLA II-restricted epitopes, we used IEDB PPC tool [41] to compute PPC for the world population using the epitope-HLA II presentation profiles predicted previously. We identified combinations of CD4 T cell epitopes reaching a maximum PPC by introducing into the IEDB PPC tool different combinations of epitopes with their corresponding HLA II binding profiles.

2.5. B Cell Epitope Prediction and Calculation of Flexibility and Solvent Accessibility. We considered flexible protein fragments identified in available 3D structures of the relevant antigens with relative solvent accessibility ≥ 50% as potential B cell epitopes. As residue flexibility values, we used normalized B factors, Z_B (2):

$$Z_B = \frac{B - \mu_B}{\partial_B}, \quad (2)$$

where B is the residue B factor from the relevant PDB, μ_B is the mean of the C_α residue of B factors, and ∂_B is the standard deviation of C_α B factors. Flexible regions, potential B cell epitopes, consisted of 9 consecutive residues or more with flexibility equal or greater than the computed ∂_B (1.0). For each selected protein fragment, we obtained a flexibility score consisting of the average flexibility of the fragment residues and a solvent accessibility value consisting of the average relative solvent accessibility (RSA) of the residues. We obtained residue RSAs from the relevant PDB coordinates using NACCESS [42]. Solvent accessibility values and flexibility scores were computed in the same manner for experimental B cell epitopes.

2.6. Blast Searches, Protein Annotation, and Analysis Procedures. We mapped epitopes onto three-dimensional (3D) structures and retrieved UniProtKB [43] entries upon BLAST searches [44] against the PDB and Swissprot databases at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). We also carried out BLAST searches with conserved epitope sequences as query against human proteins and human microbiome proteins to detect epitope identity to human or human microbiome proteins. These BLAST searches were carried out locally with standalone programs using an expectation value ($-e$) of 10,000. Human microbiome protein sequences for BLAST searches were obtained from the NIH Human Microbiome Project [45] at NCBI (https://www.ncbi.nlm.nih.gov/bioproject/43021). As human protein sequences, we used all human proteins available in the nonredundant (NR) collection at NCBI. We used PyMOL Molecular Graphics System, Version 1.8 Schrödinger, to visualize B cell epitopes on 3D structures. We identified function, subcellular localization, and temporal expression of selected EBV proteins (developmental stage) from UniProtKB [43].

3. Results

3.1. Reference EBV Proteome with Variable Residues Masked. Epitope-based vaccines can force the immune system to recognize conserved antigen regions. Therefore, a key step in our approach to epitope vaccine design is to carry out sequence variability analyses enabling the selection of conserved epitopes. To that end, we clustered all available EBV protein sequences around a reference EBV proteome (NC_007605), obtaining 85 protein clusters with EBV reference proteins on them (details in Materials and Methods). Upon aligning the sequences in the clusters, we subjected them to sequence variability analyses using the Shannon entropy, H, as variability metric. As a result, we identified that only 960 residue sites of the 42,998 evaluated had $H \geq 0.5$ and generated reference consensus EBV sequences with those

Table 1: Conserved EBV-specific CD8 T cell epitopes from early antigens.

Epitope	Antigen gene	AN[1]	HLA I restriction[2]	Predicted HLA I profile	PPC%[3]
RPPIFIRRL	EBNA3	P12977	B*07, B*08, B*0702	B*0702, B*0801, B*3501, B*5101, B*5102, B*5103, B*5301, B*5401, C*0102	57.84
SVRDRLARL	EBNA3	P12977	A*0201	A*0201, A*0203, A*0206, A*0214, B*0702, B*0801, B*1517	56.66
YVLDHLIVV	BRLF1	P03209	A*0201, A*02	A*0201, A*0202, A*0203, A*0204, A*0205, A*0206, A*0209, A*0214, B*1517, B*5701	47.34
QPRAPIRPI	EBNA6	P03204	B*0702	B*0702, B*3501, B*5101, B*5102, B*5103, B*5301, B*5401, B*5502	43.56
LPCVLWPVL	BZLF1	P03206	B*0702	B*0702, B*3501, B*5101, B*5102, B*5103, B*5301, B5401	42.4
RVRAYTYSK	BRLF1	P03209	A*0301, A*03	A*0301, A*1101, A*3101, A*3301, A*6801	41.46
AYSSWMYSY	EBNA3	P12977	A*30	A*0101, B*2701, C*0702	36.38
VLKDAIKDL	EBNA1	P03211	A*0203	A*0203, A*0204, A*0205, A*0206, A*0207, A*0214, B*0801, C*0304	33.72
QAKWRLQTL	EBNA3	P12977	B*08	B*0702, B*0801, B*1400, C*0102	32.48
RRIYDLIEL	EBNA6	P03204	B*2705	B*1400, B*2702, B*2703, B*2704, B*2705, B*2706, B*2709, C*0702	30.42
RLRAEAQVK	EBNA3	P12977	A*03, A*0301	A*0301, A*1101, B*1513	28.7
CYDHAQTHL	BMRF1	P03191	A*2402	A*0207, A*2402, B*3801	27.3
SENDRLRLL	BZLF1	P03206	B*4002, B60	B*4002, B4402	14.18
YRSGIIAVV	BMRF1	P03191	B*3906, Cw6	A*0202, A*0203, A*0204, A*0205, A*0209, B*1509, B*1510, B*1516, B*2709, B*3801, B*39,011, B*3909	12.82
ARYAYYLQF	DBP	P03227	B*2705	B*1400, B*1517, B*2701, B*2702, B*2703, B*2704, B*2705, B*2706, B*2709	7.56
VSFIEFVGW	EBNA3	P12977	B*58	B*5701, B*5702	5.08

[1]Antigen accession number from the UniProtKB database. [2]Experimental restriction profile obtained from epitope databases. [3]Average population protection coverage (PPC) of PPCs computed for 5 ethnic groups in the USA population (Black, Caucasian, Hispanic, North American natives, and Asians) using the relevant HLA I genetic frequencies [40]. The combination that reached the largest PPC (97.1%) included the CD8 T cell epitopes YVLDHLIVV, YRSGIIAVV, SVRDRLARL, RVRAYTYSK, LPCVLWPVL, and RRIYDLIEL.

variable sites masked. A variability of $H < 0.5$ is a very stringent threshold for low variability and that only a few sites (960 residue sites) with $H \geq 0.5$ were found indicates that EBV, as most dsDNA viruses, has a low mutation rate [1]. By matching EBV epitopes with this reference EBV proteome, we were able to select only those epitopes consisting of conserved residues ($H < 0.5$).

3.2. CD8 T Cell Epitope Component. To design the CD8 T cell vaccine component, we started with 88 unique EBV-specific CD8 T cell epitope sequences that were experimentally verified to be recognized in the course of a natural infection by EBV in humans. That set was reduced to 58 epitopes when we selected only those with a length of 9 residues (9 mers). We selected 9 mer peptides because most peptides presented by MHC I molecules are of that size [36]. Among those, we found 40 epitopes that did not have a single residue with $H \geq 0.5$ and none were 100% identical to human proteins or human microbiome proteins (sequences and identity data included in Additional File S2, Table S2A). A strong CD8 T cell response to early antigens is key to clear the virus [14]. Therefore, after identifying the function and developmental stage of the relevant antigens in UniprotKB, we selected 16 CD8 T cell epitopes that were present in early antigens and had a reported functionality in primary EBV infection (Table 1). For each selected CD8

Table 2: Conserved EBV-specific CD4 T cell epitopes.

Epitope	Antigen gene	AN[1]	HLA II restriction[2]	Predicted HLA II profile	PPC[3]
MLGQDDFIKFKSPLV	BFRF1	P03185	DRB1*0701	DRB1*0901, DRB1*1501, DRB1*0701, DRB1*0405, DRB1*0101, DRB1*0301, DRB5*0101, DRB1*0401	69.85
*AGLTLSLLVIC*SYLFISRG	BHRF1	P03182	DR2	DRB1*1501, DRB5*0101, DRB1*1101, DRB1*0405, DRB1*0401, DRB1*0301, DRB1*1201, DRB1*0802	57.97
LEKQLFYYIGTMLPNTRPHS	BXLF2	P03231	DR51	DRB5*0101, DRB1*1101, DRB1*0401, DRB1*0405, DRB1*1201, DRB1*1501, DRB1*0301, DRB1*0802	57.97
SRRFSWTLFL*AGLTLSLLVI*	BHRF1	P03182	DR2	DRB1*0401, DRB1*0101, DRB1*0901, DRB1*0301, DRB1*0701, DRB1*1201	55.25
SRDELLHTRAASLLY	BARF1	P0CAP6	DRB1*0701	DRB1*0701, DRB1*0101, DRB1*1201, DRB3*0202, DRB1*0901, DRB1*1302 DRB5*0101	42.9
PPVVRMFMRERQLPQ	EBNA6	P03204	HLA class II	DRB1*1101, DRB5*0101, DRB1*0301, DRB1*0401, DRB4*0101	36.88
QQRPVMFVSRVPAKK	EBNA6	P03204	HLA class II	DRB5*0101, DRB1*0802, DRB1*1101, DRB1*0301	29.35
PAQPPPGVINDQQLHHLPSG	EBNA2	P12978	DRB1*0301	DRB1*0301, DRB4 *0101	17.84
VKLTMEYDDKVSKSH	BMRF1	P03191	DRB1*0301	DRB1*0301	17.84
QKRAAPPTVSPSDTG	EBNA6	P03204	HLA class II	—	0

[1]Antigen accession number from the UniProtKB database. [2]Experimental HLA II restriction profile obtained from epitope databases. [3]Population protection coverage (PPC) was computed for the world population using the IEDB Analysis Resources tool with the HLA-DR allele reference set provided by the tool [37]. The italicized sequence is shared by the two epitopes that contain it.

T cell epitope, we predicted its potential HLA I presentation profile (see Materials and Methods) and subsequently computed the population protection coverage (PPC) for 5 distinct ethnic groups present in the USA population (see Materials and Methods). PPC of CD8 T cell epitopes ranged from 5.08% to 57.84% (Table 1). Epitopes ARYAYYLQF and VSFIEFVGW had little PPC and were discarded for further analysis. Subsequently, we used EPISOPT [39] to identify epitope combinations within the remaining 14 CD8 T cell epitopes that could provide a PPC of 95% in each one of the ethnic groups. We found that just 5 epitopes were required to reach it. Moreover, we identified 40 different epitope combinations, 3 with 5 epitopes and 37 with 6 epitopes, that reached PPC $\geq$ 95% (data not shown). EPISOPT did not report more numerous epitope combinations because adding more epitope sequences did not increase the PPC [39]. The combination with only 5 epitopes that reached the largest PPC (96.0%) consisted of epitopes YVLDHLIVV, VLKDAIKDL, RVRAYTYSK, LPCVLWPVL, and AYSSWMYSY. However, the epitope combination that provided the highest PPC (97.1%) included 6 CD8 T cell epitopes: YVLDHLIVV, YRSGIIAVV, SVRDRLARL, RVRAYTYSK, LPCVLWPVL, and RRIYDLIEL. All the 14 CD8 T cell epitopes were found in at least one of the epitope combinations reaching 95% PPC. Subsequently, we considered all the 14 CD8 T cell epitopes for inclusion in the CD8 T cell vaccine component. The selected epitopes originate from 6 different viral antigens, including EBNA3, BRLF1, EBNA6, EBNA1, BMRF1, and BZLF1 (Table 1), and thus will also contribute to a multiantigenic response.

3.3. CD4 T Cell Epitope Component. We identified a total of 21 EBV-specific CD4 T cell epitopes from the relevant epitope databases that were elicited in the course of a natural infection by EBV in humans (Table S1B in Additional File S1). Of those, we selected 10 epitopes that were conserved (Table 2) and none were 100% identical to human proteins or human microbiome proteins (see Table S2B in Additional File 2). The size of the conserved CD4 T cell peptides ranged from 15 to 20 residues long. We next identified their HLA II presentation profile by predicting peptide-MHC II binding to 15 distinct HLA-DR molecules that are frequently expressed in the population (see Materials and Methods). We chose to target HLA-DR molecules for two reasons: the alpha chain is nonpolymorphic [32] and HLA-DR are expressed at a much higher density in the cell surface of antigen-presenting cells than any other HLA II molecules [46] and thus are more relevant for epitope vaccine design [47].

Upon determining epitope HLA II presentation profiles, we computed the PPC for the world population as indicated in Materials and Methods. The maximum PPC that could be reached by considering the entire set of HLA-DR molecules is 81.81%. The PPC of selected CD4 T cell epitopes ranged from 0% (QKRAAPPTVSPSDTG) to 69.85% (MLGQDDFIKFKSPLV). The PPC that could be reached by combining all distinct HLA-DR molecules that were found to bind the selected CD4 T cell epitopes was 81.81% (Table 2). This PPC was reached by considering only the epitopes MLGQDDFIKFKSPLV, AGLTLSLLVICSYLFISRG, SRDELLHTRAASLLY, and PPVVRMFMRERQLPQ derived from antigens BFRF1, BHRF1, BARF1 and EBNA6, respectively. Antigens BFRF1 and EBNA6 are nuclear proteins, whereas BARF1 is a secreted protein and BHRF1 is a membrane-bound antigen. We considered this 4-epitope combination as the optimal CD4 T cell vaccine component.

3.4. B Cell Epitope Component. We assembled the B cell epitope vaccine component from a set of 247 EBV-specific

Table 3: Experimentally defined conserved EBV-specific B cell epitopes.

Epitope	Antigen (gene)	AN[1]	Epitope location	PDB hit[2]	Flexibility[3]	Access.[4]
SK*APESTTTSPTLNTTGFA*	gp350 (BLLF1)	P03200	Ectodomain	2H6O [422–440]	2.486 (2.672)	59.2 (63.4)
YVFYS*GNGPKASGGD*YCIQS	gp350 (BLLF1)	P03200	Ectodomain	2H6O [282–301]	1.102 (2.004)	31.7 (51.2)
QNPVYLIP*ETVPYIKWDN*	gp350 (BLLF1)	P03200	Ectodomain	2H6O [147–164]	0.618 (1.191)	62.4 (77.5)
SVKTEMLGNEID	gp350 (BLLF1)	P03200	Ectodomain	2H6O [197–208]	−0.347	19.8
QVSLESVDVYFQDVFGTMWC	gp350 (BLLF1)	P03200	Ectodomain	2H6O [122–141]	−0.575	17.5
TNTTDITYVGD	gp350 (BLLF1)	P03200	Ectodomain	2H6O [317–327]	0.121	60.1
PSTSSKLRPRWTFTSPPVTT	gp350 (BLLF1)	P03200	Ectodomain	No	N/A	N/A
QKRAAQRAAGPSVAS	gpB (BALF4)	P03188	Inner domain	No	N/A	N/A
VSGFISFFKNPFGGM	gpB (BALF4)	P03188	Transmembrane	No	N/A	N/A

[1]Accession number from UniProtKB database. [2]Epitope hit with corresponding PDBs (in bracket sequence hit). Values of [3]flexibility (arbitrary units) and [4]solvent accessibility (%) were calculated as explained in Materials and Methods. N/A: not applicable; gp: glycoprotein. We show the italicized regions in B cell epitopes consisting of 9 or more consecutive residues with flexibility ≥ 1 and we show in brackets the corresponding flexibility and accessibility values of these regions.

unique linear B cell epitope sequences ranging from 4 to 38 amino acids (Table S1C in Additional File S1). From those, we discarded B cell epitopes shorter than 9 residues and kept 117 that were conserved with no single residue with $H > 0.5$ (details in Materials and Methods). Moreover, none of these 117 B cell epitopes were identical to human proteins or to human microbiome proteins (data provided in Additional File S2, Table S2C). We analyzed the subcellular location of selected antigens to identify those that are expressed in the viral surface, accessible for antibody recognition. We found that the vast majority of the selected epitopes originated from viral intracellular antigens and therefore have no interest for B cell epitope vaccine design. We only found 9 B cell epitopes that were present in viral envelope glycoproteins: 7 from the major surface antigen gp350, the main viral determinant mediating viral attachment to B cells [48] and 2 from the envelope glycoprotein B (gB), key for the fusion of viral and host cell membranes during viral entry [49] (Table 3). However, only the 7 gp350 B cell epitopes mapped on the protein ectodomain and were further considered for the B cell epitope vaccine component. The 2 gB epitopes, QKRAAQRAAGPSVAS and VSGFISFFKNPFGGM, mapped onto the inner and transmembrane regions, respectively (Table 3).

Flexible and accessible linear B cell epitopes are often cross-reactive with antibodies against native antigens and are thereby of prime interest for epitope vaccine design [50]. Therefore, to further analyze the suitability for vaccine design of the 7 remaining gp350 B cell epitopes, we devised a system to quantify the flexibility and solvent accessibility of B cell epitopes from the known 3D structures. Briefly, we used normalized B factors and relative residue solvent accessibility computed from the relevant PDBs as measures of flexibility and accessibility (details in Materials and Methods). Following these criteria, we discarded the gp350 B cell epitope PSTSSKLRPRWTFTSPPVTT, for it mapped onto a region of the gp350 without a 3D structure and we could not readily evaluate its flexibility and accessibility. Of the 6 gp350 B cell epitopes that mapped onto the available gp350 3D structure (PDB: 2H6O), only 3 of them, SKAPESTTTSPTLNTTGFA, YVFYSGNGPKASGG DYCIQS, and QNPVYLIPETVPYIKWDN, had flexibility and solvent accessibility values supporting that they were readily accessible for antibody recognition (Table 3). In fact, visual inspection of epitopes SVKTEMLGNEID and QVSLESVDVYFQDVFGTMWC in the gp350 3D structure revealed that they were buried and thus not accessible for antibody recognition, while B cell epitope TNTTDI TYVGD though accessible (60%) was located in a rigid region of the protein (Figure S1 in Additional File S3). These epitopes will likely induce antibodies that will be unable to recognize native antigens and were discarded from the B cell vaccine component.

Following the hypothesis that highly flexible protein regions are suitable B cell epitopes for epitope vaccine design, we identified inner antigenic regions in the gp350 B cell epitopes SKAPESTTTSPTLNTTGFA, YVFYSGNGPKASG GDYCIQS, and QNPVYLIPETVPYIKWDN (APESTTTSP TLNTTGFA, GNGPKASGGD, and ETVPYIKWDN, resp.), encompassing only residues with a high degree of flexibility (≥1.0) and solvent accessibility greater than 50% (Table 3). Visual inspection of the gp350 B cell epitopes in the 3D structure clearly showed that the selected core fragments were located in highly flexibly and accessible regions of the structure while some parts of the remaining epitope were buried or semiburied (Figure 1). Therefore, we regarded the antigenic core regions (APESTTTSPTLNTTGFA, GNGPKA SGGD, and ETVPYIKWDN) identified in the gp350 B cell epitopes as the experimental B cell component of the EBV epitope vaccine ensemble.

As all experimental B cell epitopes suitable for epitope vaccine design were in gp350, we sought to identify potential B cell epitopes from the 3D structures of EBV envelope proteins gp42 (PDB: 3FD4), gB (PDB: 3FVC) and the heterodimer conformed by gH and gL (PDB: 5T1D). These proteins have been described to participate in the viral attachment and/or fusion to the host cell membrane required for viral entry [49, 51, 52]. We considered as potential B cell epitopes, antigen fragments in the relevant 3D structures consisting of 9 or more consecutive residues with flexibility ≥ 1.0 and an average accessibility ≥ 50% (details in Materials and

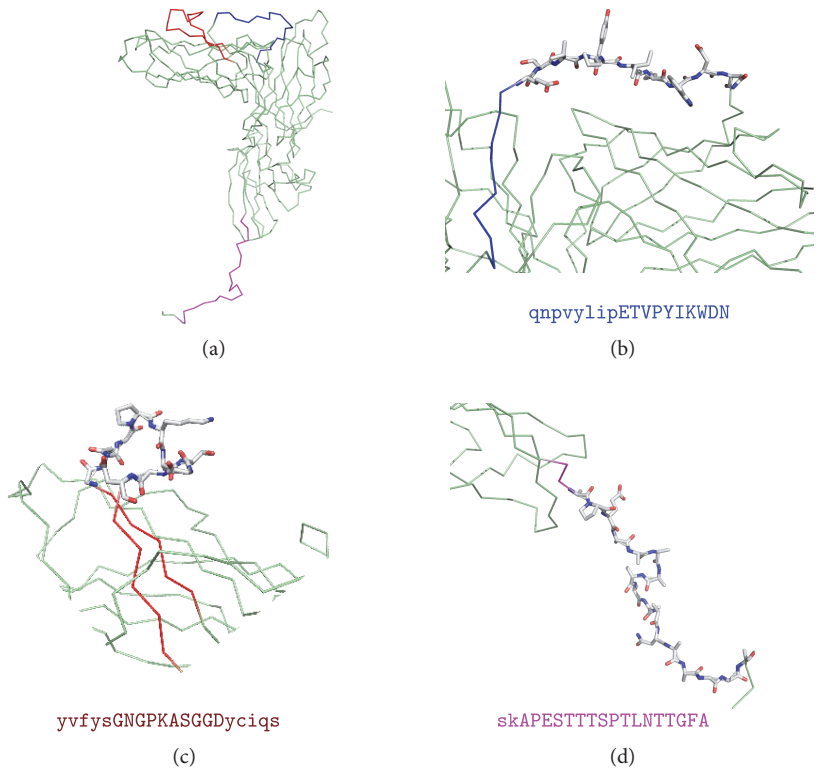


Figure 1: Structural mapping of selected experimental EBV-specific B cell epitopes. Conserved EBV epitopes map onto two different regions of the 3D structure of gp350 (PDB code: 2H6O): QVNYLIPETVPYIKWDN and YVFYSGNGPKASGGDYCIQS map at the glycan-free surface of the CR2 receptor binding site; SKAPESTTTSPTLNTTGFA maps at the C-term tail of the PDB. (a) General view of gp350 featured as ribbon with B cell epitopes highlighted in red, blue, and purple. Protein regions of the selected epitopes are zoomed in panels (b, c, d). We show in sticks the part of the epitopes that exhibited greater flexibility and accessibility which was ultimately selected for the proposed vaccine ensemble. In ribbon, we show the B cell epitope residues that do not comply with the flexibility and accessibility criteria (typed in a minor case in the corresponding sequence indicated at the bottom of each panel). Figures were rendered using PyMOL.

Table 4: Predicted conserved B cell epitopes from EBV envelope proteins.

Epitope	Antigen (gene)	Accession number[1]	PDB[2]	Flex.[3]	Acc. (%)[4]	BLAST hit HMP (%)[5]	BLAST hit human (%)[6]
KLPHWTPTLH	gp42 (BZLF2)	P03205	3FD4:A [45–54]	2.256	80.0	EJZ65106.1 (70.00)	AAH22472.1 (60.00)
NTTVGIELPDA	gpB (BALF4)	P03188	3FVC [307–317]	1.890	67.0	EHM53795.1 (72.73)	XP_011519547.1 (63.64)
SSHGDLFRFSSDIQCP	gpB (BALF4)	P03188	3FVC [32–47]	1.369	69.8	KGF26221.1 (50.00)	XP_011520599.1 (50.00)
FSVEDLFGAN	gL (BKRF2)	P03212	5T1D:B [95–104]	1.505	53.1	EKB09257.1 (65.00)	XP_005271219.1 (70.00)

[1]Accession number from the UniProtKB database. [2]Epitope location in their corresponding PDBs is shown in brackets. The specific chain is indicated along with the PDB code. [3]Values of flexibility (arbitrary units) and [4]solvent accessibility (%) were calculated as explained in Materials and Methods. [5,6]Accession number of closest epitope BLAST hit in human microbiome proteins and human proteins, respectively (percentage of identity in parenthesis).

Methods). As a result, we identified a potential B cell epitope in gp42 protein (KLPHWTPTLH), two at the gB protein (NTTVGIELPDA and SSHGDLFRFSSDIQCP), and one in the gL monomer (FSVEDLFGAN) (Table 4). No epitopes fulfilling the required criteria were identified at the gH protein. These predicted B cell epitopes were mapped to their corresponding 3D structures to confirm that they were in readily accessible regions for antibody recognition (Figure 2). KLPHWTPTLH mapped at the N-terminal region of gp42, which is involved in gH interaction and sits opposite to the HLA-DR binding site of the molecule (colored in red in Figure 2(a)). The gB epitopes mapped onto two distinct regions, domains II and III, that are likely relevant for interaction with other glycoproteins involved in viral entry [49] (Figures 2(a) and 2(b)). The single gL epitope mapped in a region in close proximity to gH and the binding site of a monoclonal antibody (mAb) EID1 that interferes with EBV infection of

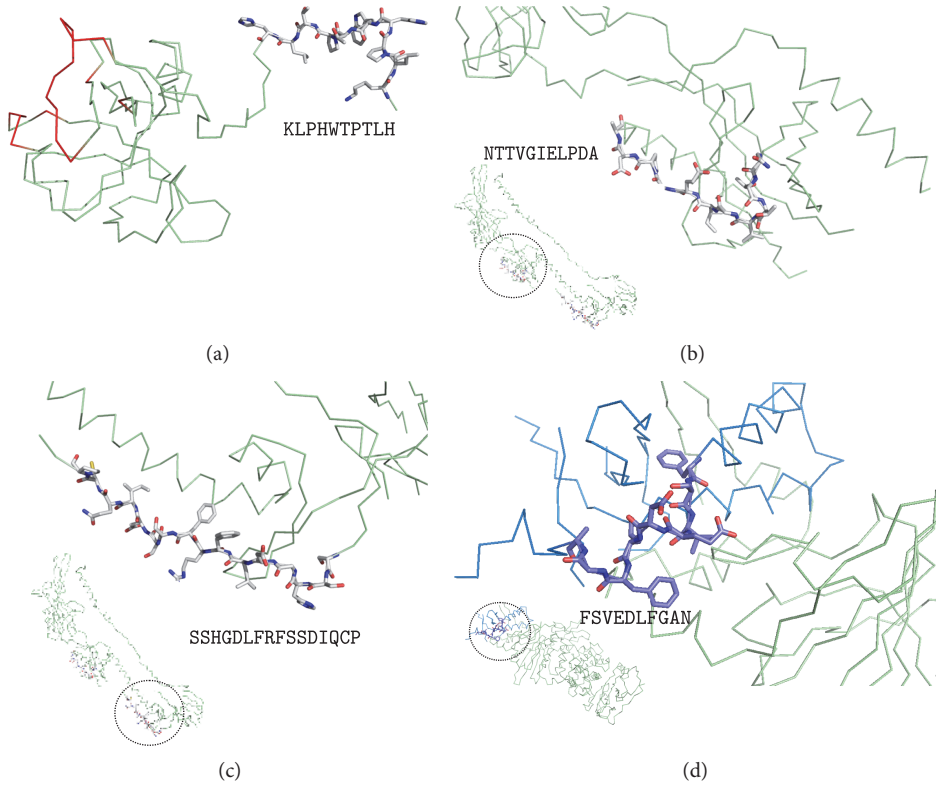


Figure 2: Structural mapping of predicted B cell epitopes in EBV envelope proteins. (a) KLPHWTPTLH in EBV gp42 3D structure (PDB: 3FD4 chain A); epitope shown as sticks and gp42 region interaction with HLA-DR is shown in red. (b) NTTVGIELPDA and (c) SSHGDLFRFSSDIQCP at EBV gB 3D structure (PDB: 3FVC) map, respectively, in its domain II and domain III; epitopes shown as sticks. (d) FSVEDLFGAN at gL 3D structure (PDB: 5T1D chain B) in its domain I (colored in blue); gH is colored in pale green. In (b, c, d), the corresponding whole structure is shown minimized at the bottom left of each panel; the magnified epitope mapping region is circled in them. In (a, b, c), the protein backbone is featured as pale green ribbon. Figures were rendered using PyMOL.

epithelial cells [52] (Figure 2(d)). We also verified that none of the predicted B cell epitopes were identical to human proteins or human microbiome proteins (Table 4).

4. Discussion

Over 90% of human adults are infected with EBV. Most infections occur in childhood and are asymptomatic or course with nonspecific symptoms. Nonetheless, EBV is the primary cause of IM when infection occurs in early adulthood. Furthermore, the viral infection is associated with autoimmunity and a number of lymphocyte and epithelial cell malignancies [18, 19]. Despite its wide impact, there is no treatment available, hence the growing interest in finding a prophylactic and/or therapeutic EBV vaccine.

The target population for an EBV prophylactic vaccine in the developed world would be 10- or 11-year-old children, before they are susceptible to most severe IM symptomatologies. It is acknowledged that by precluding the initial viral infection, the risks of developing EBV-associated autoimmune and cancer disorders would also be reduced [53]. In sub-Saharan Africa and southern China, where Burkitt's lymphoma and nasopharyngeal carcinoma are major public health problems and children are infected by EBV earlier in life, the vaccine target would be much younger infants. EBV-naïve transplant recipients susceptible to suffer posttransplant lymphoproliferative disorders (PTLD) would also benefit from a prophylactic vaccine [54].

Currently, the most advanced EBV vaccine clinically tested consists of a gp350 subunit that was administered with AS04 adjuvant to virus-naïve young adults [55]. The gp350 subunit vaccination strategy follows the approach successfully used in other viral infections, that is, induction of neutralizing antibodies (nAbs) against the most abundant glycoprotein on the virus, which also represents the main target of naturally occurring nAbs [16]. In this regards, a microneutralization assay based on an EBV expressing green-fluorescent protein has been very recently developed to provide measurement of humoral EBV vaccine responses in large clinical trials [56]. Another EBV vaccine trial was designed to control the expansion of EBV-infected B cells, based on the generation of CD8 T cell immunity to EBNAs [57]. Specifically, the vaccine consisted of a single EBNA3A epitope restricted by HLA-B08 administered as a peptide along with tetanus toxoid as adjuvant [57].

A major outcome of the Sokal et al. [55] clinical trial was that immunization with gp350 did not protect from new viral infections [55]. Therefore, it has been suggested that a prophylactic vaccine against EBV should elicit B cell responses also against all 5 major viral envelope proteins involved in

host-cell attachment and entry, including gp42, gH, gL, BMRF2 (gp350), and gB [58]. Among these, at least the first four are known to elicit neutralizing antibodies [59]. The induction of cytotoxic T cell responses against early viral antigens has been as well suggested in order to destroy recently infected B cells [14, 53, 59]. Attaching to these premises, we used a computer-assisted strategy to design a prophylactic epitope vaccine ensemble against EBV infection.

The strategy that we followed to design the EBV vaccine relied on combining legacy experimentation consisting of experimentally defined epitopes with immunoinformatics predictions. This strategy was first conceived to assemble CD8 T cell epitope vaccines [21, 39] and latter extended to include CD4 T cell epitope vaccines [22]. The main advantage of this approach is that of saving time and resources as it mainly relies on experimentally validated epitopes, not on predicted epitopes, using immunoinformatics to identify those that are more suitable for epitope vaccine design. It is worth noting that epitope prediction is not a precise science and epitope prediction methods only facilitate epitope discovery by providing candidates that need to be validated experimentally. Therefore, our strategy ought to gain widening acceptance as a vaccine design tool whenever ample experimental epitope data is readily available. Key criteria for epitope inclusion/selection are conservation and binding to multiple MHC molecules for maximum population protection coverage. Here, we also added that the source of CD8 T cell epitopes had to be from early EBV antigens with defined function in the primary infection process. Moreover, we checked that peptides were nonself and did not have exact matches with human proteins or human microbiome proteins and extended the approach to B cell epitopes. To that end, we devised a system to select from experimentally defined B cell epitopes those that were conserved, nonself and located on the ectodomains of viral envelope antigens and consisted of highly flexible and solvent-accessible residues (Figure 3). Note that we are not discriminating B cell epitopes from non-B cell epitopes in primary sequences. In fact, solvent accessibility or flexibility alone cannot discriminate B cell epitopes from non-B cell epitopes in primary sequences [60]. Instead, we are selecting known B cell mapping in the antigen surface that isolated from the antigen context can elicit antibodies cross-reacting with the native antigens and hence are worth for epitope vaccine design [61].

The composition of the epitope vaccine ensemble designed in this study includes 14 CD8 T cell epitopes, 4 CD4 T cell epitopes, and 7 B cell epitopes (Table 5). None of these epitopes matched exactly to human proteins or human microbiome proteins. This result is somewhat predictable for we focused mostly on epitopes that have been verified experimentally and it should be expected that the immune system selected nonself targets for recognition. Nonetheless, a few of the selected epitopes have a high identity with human microbiome proteins (around 88.9%, Table 5). Whether this high identity to human microbiome proteins could be a source of trouble is arguable: detection of epitope identity to self-proteins required using BLAST

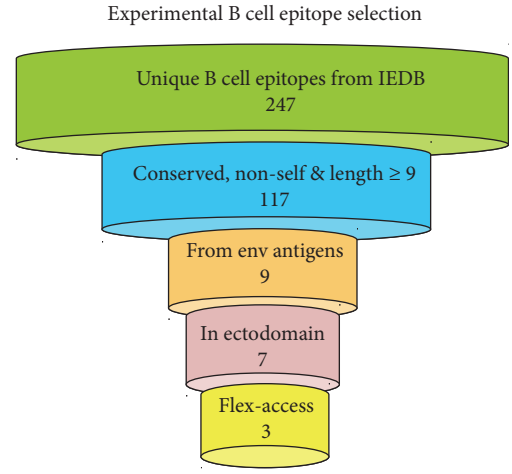


Figure 3: Strategy for experimental B cell epitope selection. Overview of the approach devised to select invariant experimental EBV-specific B cell epitopes for the B cell component of an epitope-based vaccine against EBV. The approach comprises 5 steps: (1) selection of unique epitopes from databases; (2) sequence variability filtering and testing for self-peptides; (3) selection of epitopes from viral envelope antigens; (4) progression of epitopes located to envelope protein ectodomains; (5) final output of epitopes that fulfill the flexibility and accessibility criteria established in the text. None of the epitopes that we selected were identical to human proteins or proteins from the human microbiome.

with expectation values of 10000, epitope matches may not be available for recognition, and epitope recognition can be disrupted by single amino acid changes.

According to some authors, the ideal EBV CD8 T cell epitope component should include antigens EBNA2, EBNA-LP, and BHRF1, which are abundant at the very initial stage of B cell infection [14]. Our epitope vaccine ensemble does not include CD8 T cell epitopes from these three antigens. However, it includes CD8 T cell epitopes from other EBV early antigens, such as EBNA1, EBNA3, EBNA6, BMRF1, BRLF1, and BZLF1 (Tables 1 and 5). Although a 95% PPC was reached with just 5 CD8 T cell epitopes, the key importance of a broad multiantigenic cytotoxic response prompted us to incorporate 14 CD8 T cell epitopes. For the CD4 T cell component, our proposed vaccine ensemble includes 4 epitopes reaching the maximum PPC possible of 81.8% provided by the reference set of HLA II molecules targeted for binding predictions [37]. The PPC of the CD4 T cell component is likely an underestimation. HLA II molecules are very promiscuous [62] and the selected epitopes will surely bind and be presented by other HLA II molecules not included in the selected reference set [37].

For the B cell epitope vaccine component, we included 7 B cell epitopes consisting of 3 experimental B cell epitopes from gp350 plus 4 other predicted B cell epitopes from EBV envelope proteins gp42, gB, and gL, all of them continuous and with high flexibility and solvent accessibility. We focused on linear B cell epitopes because they can be delivered isolated from their antigen context to induce selective humoral responses. We sought to predict B cell epitopes on gp42, gB, and gL that can be used to elicit antibodies that

Table 5: Proposed epitope vaccine ensemble for EBV.

CD8 T cell epitope vaccine component					
Sequence	Antigen	AN[1]	BLAST hit HMP[2]	BLAST hit humans[3]	PPC%[4]
RPPIFIRRL	EBNA3	P12977	EFI49553.1 (55.56)	NP_001182344.1 (66.67)	
SVRDRLARL	EBNA3	P12977	No hit (—)	3HR0 (55.56)	
YVLDHLIVV	BRLF1	P03209	EPH07203.1 (88.89)	XP_011535331.1 (66.67)	
QPRAPIRPI	EBNA6	P03204	EEZ70880.1 (66.67)	AFC01212.1 (55.56)	
LPCVLWPVL	BZLF1	P03206	ETN46892.1 (77.78)	XP_011511695.1 (55.56)	
RVRAYTYSK	BRLF1	P03209	EEY91922.1 (66.67)	CAE46202.1 (55.56)	>95
AYSSWMYSY	EBNA3	P12977	EKB85112.1 (77.78)	EAW88404.1 (66.67)	
VLKDAIKDL	EBNA1	P03211	KXB56071.1 (88.89)	EAX00446.1 (66.67)	
QAKWRLQTL	EBNA3	P12977	EHR35488.1 (66.67)	XP_005255827.1 (77.78)	
RRIYDLIEL	EBNA6	P03204	EDS12420.1 (77.78)	EAW88480.1 (66.67)	
RLRAEAQVK	EBNA3	P12977	No hit (—)	XP_011507142.1 (77.78)	
CYDHAQTHL	BMRF1	P03191	EFF75621.1 (77.78)	CAH10644.1 (66.67)	
SENDRLRLL	BZLF1	P03206	EGG37664.1 (77.78)	EAW88969.1 (77.78)	
YRSGIIAVV	BMRF1	P03191	OFQ99895.1 (88.89)	BAC03504.1 (66.67)	
CD4T cell epitope vaccine component					
Sequence	Antigen	AN[1]	BLAST hit HMP[2]	BLAST hit humans[3]	PPC%[4]
MLGQDDFIKFKSPLV	BFRF1	P03185	EIY33207.1 (46.67)	NP_001284364.1 (53.33)	
AGLTLSLLVICSYLFISRG	BHRF1	P03182	EKN19533.1 (47.37)	EAW92092.1 (52.63)	>81.8
SRDELLHTRAASLLY	BARF1	P0CAP6	EPB87510.1 (66.67)	XP_011514101.1 (66.67)	
PPVVRMFMRERQLPQ	EBNA6	P03204	EFV04068.1 (46.67)	AAP34452.1 (60.00)	
B cell epitope vaccine component					
Sequence	Antigen	AN[1]	BLAST hit HMP[2]	BLAST hit humans[3]	Src.[5]
APESTTTSPTLNTTGFA	gp350 (BLLF1)	P03200	EGY79509.1 (58.82)	NP_001276932.1 (52.94)	E
GNGPKASGGD	gp350 (BLLF1)	P03200	EHM51909.1 (70.00)	NP_055501.2 (70.00)	E
ETVPYIKWDN	gp350 (BLLF1)	P03200	EET62946.1 (50.00)	NP_001193968.1 (50.00)	E
KLPHWTPTLH	gp42 (BZLF2)	P03205	EJZ65106.1 (70.00)	AAH22472.1 (60.00)	P
NTTVGIELPDA	gpB (BALF4)	P03188	EHM53795.1 (72.73)	XP_011519547.1 (63.64)	P
SSHGDLFRFSSDIQCP	gpB (BALF4)	P03188	KGF26221.1 (50.00)	XP_011520599.1 (50.00)	P
FSVEDLFGAN	gL (BKRF2)	P03212	EKB09257.1 (80.00)	XP_005271219.1 (70.00)	P

[1]Accession number from UniProtKB database. [2,3]Accession number of the closest epitope BLAST hit to human microbiome proteins and human proteins, respectively (percentage of identity in parenthesis). [4]Population protection coverage (PPC) of the CD8 and CD4 T cell epitope ensemble. [5]Src., source, whether the epitope derived from an experimental B cell epitope (E) or it was predicted (P).

are cross-reactive with the native antigens. To that end, we needed to identify solvent-exposed B cell epitopes in the mentioned antigens and we could have used a number of methods to predict conformational B cell epitopes from the available 3D structures (reviewed in [63]). However, conformational B cell epitopes can not be isolated from their protein context and used as immunogens. Therefore, we turned our attention to linear B cell epitopes as they can be delivered isolated from the antigen and induce selective humoral responses. There are also a number of methods to predict linear B cell epitopes from primary sequences (reviewed in [60, 64]), but the predicted epitopes seldom match in solvent-accessible regions and are notoriously unreliable [60, 65, 66]. Hence, in this study, we assumed that highly flexible and solvent-accessible fragments in protein surfaces are potential linear B cell epitopes [50] and devised a system to identify them from the relevant 3D structures (details in Materials and Methods). Specifically, predicted B cell epitopes consisted of conserved fragments with at least 9 consecutive residues with flexibility (normalized B factor) > 1 and an average relative solvent-exposed accessibility ≥ 50%.

Analysis of the structural mapping of the selected B cell epitopes onto the relevant 3D structure can reveal their importance for epitope vaccine design. The gp42 B cell epitope (KLPHWTPTLH) is located in the N-terminal portion of the protein far and opposite from the HLA-DR contact region (Figure 2(a)). Therefore, antibodies against this gp42 B cell epitope will unlikely block the gp42 interaction with HLA-DR required for viral entry into B cells. The gp42 N-terminal region, where KLPHWTPTLH maps, interact with gH at a site in close proximity to the β1-integrin-binding motif "KGD" [52]. Both gp42 and peptides from the N-terminal region of gp42 that binds to gH interfere with β1-integrin interaction and viral entry in epithelial cells

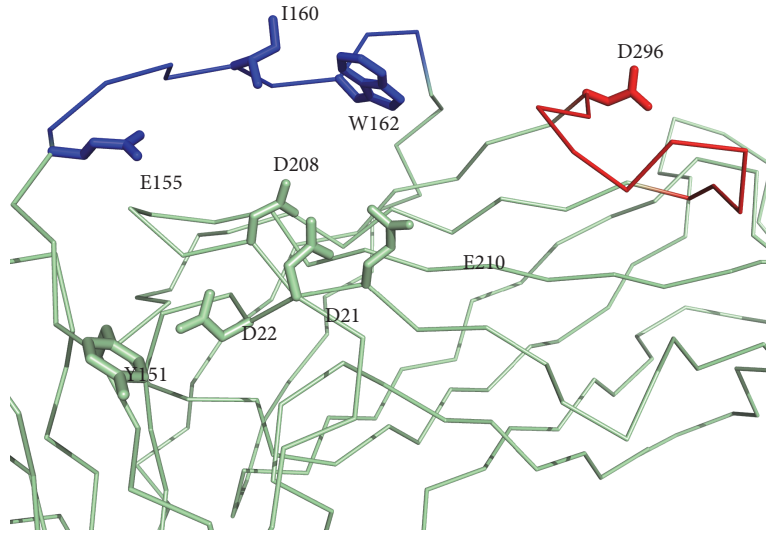


Figure 4: The EBV gp350 contact region with CR2. EBV B cell epitopes ETVPYIKWDN and GNGPKASGGD map onto a gp350 region that interacts with CR2; epitopes colored blue and red and the gp350 backbone featured as pale green ribbon. Side chains of the residues described to interact with CR2 receptor by Young et al. [67] are shown as sticks. Figure was rendered using PyMOL.

[52]. In this context, the role of antibodies against this gp42 epitope with regard to viral entry in epithelial cells is unclear. Binding of antibodies to the epitope when gp42 is in complex with gH could prevent epithelial infection by EBV. However, such prevention is unlikely if antibodies against the epitope block the interaction between gp42 and gH. Despite poor neutralizing qualities of the gp42 B cell epitope KLPHWTPTLH, antibodies against it could still contribute to viral clearance by promoting complement activation and phagocytosis. The two predicted B cell epitopes in gB, NTTVGIELPDA, and SSHGDLFRFSSDIQCP, mapped onto two distinct protein domains (Figures 2(b) and 2(c)) that are thought to be relevant in the mechanism of EBV fusion to host membranes [49]. Hence, antibodies binding at this region could interfere in the vital fusion step required for viral entry. The B cell epitope predicted in gL, FSVEDLFGAN, mapped onto a region intertwined with gH and is in close proximity to the binding site of mAb E1D1 [52]. This antibody has been described to inhibit gH fusion to epithelial cells despite locating far from the gH integrin binding site (KGD). Whether an antibody against gL-protruding epitope FSVEDLFGAN might also exert a similar distant effect is unknown but remains a possibility.

Flexibility and accessibility were also key criteria to select and refine experimental B cell epitopes, leading to the selection of the gp350 B cell epitopes ETVPYIKWDN, GNGPKASGGD, and APESTTTSPTLNTTGFA (Table 3 and Figure 1). Two of these B cell epitopes, ETVPYIKWDN and GNGPKASGGD, mapped onto the glycan-free region of gp350 described to interact with the CR2 receptor [48]. Furthermore, residues E155, I160, and W162 from ETVPYIKWDN and D296 from GNGPKASGGD have been shown to contact the CR2 receptor (Figure 4) [67]. Noteworthy, the well-characterized EBV nAb 72A1 binds to gp350 in this glycan-free region [67]. Therefore, B cell epitopes ETVPYIKWDN and GNGPKASGGD have a great potential to induce neutralizing antibodies. In fact, GNGPKASGGD and ETVPYIKWDN are within peptide fragments that have been shown already to elicit antibodies that block binding of mAb 72A1 to gp350 [68]. Lastly, epitope APESTTTSPTLNTTGFA mapped onto the C-terminal end of the solved structure of gp350 (Figure 1). Mutagenesis of its E425 and S426 residues did not inhibit binding of gp350 to mAb 72A1 [48]. Although initially far from the receptor interaction region and containing a glycosylated asparagine residue (N435), it cannot be discarded that an antibody targeting it could help to control viral infection, for example through antibody-mediated complement activation and phagocytosis. Overall, these results validate the conservancy, flexibility, and accessibility criteria followed for the selection and prediction of B cell epitopes.

We trust that the application of the knowledge-based approach depicted in this work to design an epitope vaccine ensemble against EBV can save time and effort developing such a vaccine, as most of the components consist on experimentally defined EBV-specific epitopes. However, our epitope-based vaccine ensemble is theoretical, and extra validations will be required prior to formulating a vaccine that can actually be tested. For example, T cell epitopes used in our vaccine have been shown to be immunogenic in the context of experimentally defined HLA restriction elements (see Tables 1 and 2). However, we predicted that these epitopes will be also immunogenic in the context of different HLAs. To test that, T cells from subjects expressing the relevant HLA molecules can be expanded using dendritic cells loaded with the corresponding epitope peptides and cloned. Subsequently, T cell clone immunoreactivity can be checked through a number of assays (ELISPOT, intracellular cytokine staining, etc) using B-LCL 721.221 cells expressing single HLA molecules as described elsewhere [21, 69]. Selected B

cell epitopes should also be subjected to extra validations, in particular to test whether they elicit antibodies cross-reacting with native antigens. To that end, sera from immunized mice with B cell epitope peptides could be used to check whether they recognize native antigens in ELISA assays and/or interfere with EBV infection of epithelial and B cells as described elsewhere [68]. Once the individual components of the epitope vaccine ensemble had passed experimental validation, it will still remain to elucidate how to formulate such a vaccine for delivering the epitopes.

There are several choices to formulate epitope vaccines ranging from peptide-based formulations to genetic formulations. Regardless of the choice, CD4 T cell epitopes need to be physically linked with the other selected epitopes, particularly B cell epitopes, to elicit productive Th cells [70]. A peptide-based vaccine has already been tested for the delivery of an EBV CD8 T cell epitope fused with tetanus toxoid to increase immunogenicity and elicit Th responses [57]. Similarly, a polymeric epitope concatemer in the form of a "string-of-beads" could be chemically synthesized or formulated as a genetic construct [71]. In either cases, the order of the epitopes and the presence of cleavage sites between them are crucial features to address [71]. Concatenating epitopes can result in toxic products and tools to predict toxicity can also be used to optimize epitope concatemers [72]. Toxicity of epitope vaccine formulations should nevertheless be checked in cellular assays prior to carrying out any immunization studies. In general, poor immunogenicity is an important issue with peptide-based formulations [22]. A recent development in vaccine formulation that increases the immunogenicity of the epitope-peptide components consists in the use of nanoparticles of diverse nature [73]. For example, Kuai et al. [74] used high-density lipoprotein-mimicking nanodiscs coupled with peptides to stimulate potent tumor-specific CD8 T cell responses that inhibited tumor growth in a murine model of colon carcinoma. Nanoparticles have also been used to deliver genetic constructs, particularly RNA constructs. RNA-based vaccine formulations offer lower safety concerns and enhanced immunogenicity with regard to those based on DNA, and inherent RNA instability can be overcame using nanoparticles for delivery [75].

Ideally, the B cell response should only be focused on B cell epitopes. To that end, a solution would be formulating the epitope vaccine as liposomal or virosome-like particles, where the selected T cell epitopes, either alone or concatenated, ought to be placed encapsulated inside the particle and the B cell epitopes displayed linked in the outer part of the particle [76, 77]. These liposomal vaccine formulations are also more immunogenic than those consisting of genetic or synthetic peptide-based constructs [76, 77]. Moreover, immunogenicity can be further enhanced by the inclusion of appropriated adjuvants [78].

Epitope vaccine formulations, as any vaccine candidate, should be evaluated in preclinical animal models prior to clinical testing in humans. However, in the case of EBV, this stands as a major drawback as there is a lack of appropriate animal models that recapitulate EBV infection and its immune control [79]. Thus, EBV vaccine immunogenicity and protection capabilities have to be assessed in clinical studies. Although this is very informative and may accelerate the developmental process, it also carries high associated costs early in the discovery path and involves enrollment of participants, which is not at the reach of many research groups. The clinical status of the target population to test EBV prophylactic vaccine candidates should also be considered. For instance, the phase II study by Sokal et al. [55], the most advanced of any EBV vaccine tested so far [54], involved a total of 181 EBV-seronegative, healthy, young volunteers between 16 and 25 years of age that were randomized in a double-blind fashion to receive either placebo or a recombinant EBV subunit glycoprotein 350.

5. Conclusions and Limitations

EBV infection is associated with a number of human diseases, including cancer and autoimmunity. Currently, it is unclear why some individuals with apparently proper responses to EBV develop associated diseases while others do not, but surely genetic and environmental factors, including life style and past pathogen encounters, play a role [80–82]. In any case, a prophylactic EBV vaccine will be beneficial in preventing EBV-associated diseases [53, 59]. We herein provide an epitope ensemble that would serve to develop an epitope-based prophylactic vaccine against EBV infection, eliciting both adaptive cellular and humoral immunity. The T cell component consists of highly conserved experimental EBV-specific epitopes capable of eliciting cellular responses in virtually the whole population. The B cell component consists of conserved experimental and predicted B cell epitopes from EBV envelope proteins gp350, gp42, gB, and gL. These epitopes were selected from the relevant 3D structures applying a novel structure-based reverse vaccinology approach that includes calculation of flexibility and solvent accessibility values. As a result, we identified B cell epitopes that could elicit antibodies interfering with EBV entry in epithelial and B cells. Whether our epitope vaccine ensemble has also any therapeutic value is arguable but, clearly, it is harder to combat EBV once it has established a latent infection.

This study has limitations that may handicap its translation into an EBV vaccine. Appropriate antigen processing is a key limiting factor in the immunogenicity of T cell epitopes [83]. Therefore, we selected experimental T cell epitopes that were shown to be processed and presented in the course of a natural infection with EBV and assumed that T cell epitope immunogenicity will be then only determined by their binding to MHC molecules. This assumption has not been thoroughly tested and it is very sensitive to possible errors in the databases where we collected the data. In the same line, population coverage estimates for the T cell component need to be tested as they are inferred from peptide binding predictions to MHC molecules. Nonetheless, the reliability of peptide-MHC binding predictions has been widely proved [84]. With regard to the B cell component, we deliberately failed to include conformational epitopes as they cannot be isolated from their context and solely focused on linear B cell epitopes. Whether these B cell epitopes are able to elicit

antibodies recognizing the native protein conformations needs to be tested.

Abbreviations

MHC: Major histocompatibility complex
HLA: Human leukocyte antigens
gp: Glycoprotein
nAb: Neutralizing antibody.

Conflicts of Interest

Julio Alonso-Padilla is a postdoctoral researcher at ISGlobal supported by the Juan de la Cierva Program (MINECO, Spain) and a visiting scientist at the Laboratory of Immunomedicine, Faculty of Medicine, UCM, led by Pedro A. Reche. ISGlobal is a member of the CERCA Programme, Generalitat de Catalunya. The authors declare that they have no conflict of interests.

Acknowledgments

The authors wish to thank *Inmunotek S.L.* and the Spanish Department of Science at MINECO for supporting the research of the Immunomedicine Group through Grants SAF2006:07879, SAF2009:08301, and BIO2014:54164-R to Pedro A. Reche. Julio Alonso-Padilla acknowledges the support provided by Joaquim Gascón, director of the ISGlobal Chagas Disease Program.

References

[1] Z. Lin, X. Wang, M. J. Strong et al., "Whole-genome sequencing of the Akata and Mutu Epstein-Barr virus strains," *Journal of Virology*, vol. 87, no. 2, pp. 1172–1182, 2013.

[2] K. Sathiyamoorthy, J. Jiang, Y. X. Hu et al., "Assembly and architecture of the EBV B cell entry triggering complex," *PLoS Pathogens*, vol. 10, no. 8, article e1004309, 2014.

[3] M. Neves, J. Marinho-Dias, J. Ribeiro, and H. Sousa, "Epstein-Barr virus strains and variations: geographic or disease-specific variants?," *Journal of Medical Virology*, vol. 89, no. 3, pp. 373–387, 2017.

[4] L. S. Young and A. B. Rickinson, "Epstein-Barr virus: 40 years on," *Nature Reviews Cancer*, vol. 4, no. 10, pp. 757–768, 2004.

[5] E. K. Vetsika and M. Callan, "Infectious mononucleosis and Epstein-Barr virus," *Expert Reviews in Molecular Medicine*, vol. 6, no. 23, pp. 1–16, 2004.

[6] G. C. Faulkner, S. R. Burrows, R. Khanna, D. J. Moss, A. G. Bird, and D. H. Crawford, "X-linked agammaglobulinemia patients are not infected with Epstein-Barr virus: implications for the biology of the virus," *Journal of Virology*, vol. 73, no. 2, pp. 1555–1564, 1999.

[7] C. D. Shannon-Lowe, B. Neuhierl, G. Baldwin, A. B. Rickinson, and H. J. Delecluse, "Resting B cells as a transfer vehicle for Epstein-Barr virus infection of epithelial cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 18, pp. 7065–7070, 2006.

[8] B. Kempkes and E. S. Robertson, "Epstein-Barr virus latency: current and future perspectives," *Current Opinion in Virology*, vol. 14, pp. 138–144, 2015.

[9] M. M. Mullen, K. M. Haan, R. Longnecker, and T. S. Jardetzky, "Structure of the Epstein-Barr virus gp42 protein bound to the MHC class II receptor HLA-DR1," *Molecular Cell*, vol. 9, no. 2, pp. 375–385, 2002.

[10] G. S. Taylor, H. M. Long, J. M. Brooks, A. B. Rickinson, and A. D. Hislop, "The immunology of Epstein-Barr virus-induced disease," *Annual Review of Immunology*, vol. 33, pp. 787–821, 2015.

[11] D. A. Thorley-Lawson, "Epstein-Barr virus: exploiting the immune system," *Nature Reviews Immunology*, vol. 1, no. 1, pp. 75–82, 2001.

[12] T. Strowig, F. Brilot, F. Arrey et al., "Tonsilar NK cells restrict B cell transformation by the Epstein-Barr virus via IFN-gamma," *PLoS Pathogens*, vol. 4, no. 2, article e27, 2008.

[13] A. D. Hislop, G. S. Taylor, D. Sauce, and A. B. Rickinson, "Cellular responses to viral infection in humans: lessons from Epstein-Barr virus," *Annual Review of Immunology*, vol. 25, pp. 587–617, 2007.

[14] J. M. Brooks, H. M. Long, R. J. Tierney et al., "Early T cell recognition of B cells following Epstein-Barr virus infection: identifying potential targets for prophylactic vaccination," *PLoS Pathogens*, vol. 12, no. 4, article e1005549, 2016.

[15] E. Amyes, C. Hatton, D. Montamat-Sicotte et al., "Characterization of the $CD4^+$ T cell response to Epstein-Barr virus during primary and persistent infection," *The Journal of Experimental Medicine*, vol. 198, no. 6, pp. 903–911, 2003.

[16] W. Bu, G. M. Hayes, H. Liu et al., "Kinetics of Epstein-Barr virus (EBV) neutralizing and virus-specific antibodies after primary infection with EBV," *Clinical and Vaccine Immunology*, vol. 23, no. 4, pp. 363–369, 2016.

[17] M. De Paschale and P. Clerici, "Serological diagnosis of Epstein-Barr virus infection: problems and solutions," *World Journal of Virology*, vol. 1, no. 1, pp. 31–43, 2012.

[18] M. P. Thompson and R. Kurzrock, "Epstein-Barr virus and cancer," *Clinical Cancer Research*, vol. 10, no. 3, pp. 803–821, 2004.

[19] A. Ascherio and K. L. Munger, "EBV and autoimmunity," *Current Topics in Microbiology and Immunology*, vol. 390, Part 1, pp. 365–385, 2015.

[20] J. I. Cohen, "Epstein-Barr virus vaccines," *Clinical & Translational Immunology*, vol. 4, no. 1, article e32, 2015.

[21] P. A. Reche, D. B. Keskin, R. E. Hussey, P. Ancuta, D. Gabuzda, and E. L. Reinherz, "Elicitation from virus-naive individuals of cytotoxic T lymphocytes directed against conserved HIV-1 epitopes," *Medical Immunology*, vol. 5, p. 1, 2006.

[22] Q. M. Sheikh, D. Gatherer, P. A. Reche, and D. R. Flower, "Towards the knowledge-based design of universal influenza epitope ensemble vaccines," *Bioinformatics*, vol. 32, no. 21, pp. 3233–3239, 2016.

[23] M. Molero-Abraham, J. P. Glutting, D. R. Flower, E. M. Lafuente, and P. A. Reche, "EPIPOX: immunoinformatic characterization of the shared T-cell epitome between variola virus and related pathogenic Orthopoxviruses," *Journal of Immunology Research*, vol. 2015, Article ID 738020, 11 pages, 2015.

[24] C. M. Diez-Rivero and P. A. Reche, "CD8 T cell epitope distribution in viruses reveals patterns of protein biosynthesis," *PLoS One*, vol. 7, no. 8, article e43674, 2012.

[25] P. A. Reche, H. Zhang, J. P. Glutting, and E. L. Reinherz, "EPIMHC: a curated database of MHC-binding peptides for

customized computational vaccinology," *Bioinformatics*, vol. 21, no. 9, pp. 2140-2141, 2005.

[26] Q. Zhang, P. Wang, Y. Kim et al., "Immune epitope database analysis resource (IEDB-AR)," *Nucleic Acids Research*, vol. 36, Web Server issue, pp. W513–W518, 2008.

[27] W. Li and A. Godzik, "Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences," *Bioinformatics*, vol. 22, no. 13, pp. 1658-1659, 2006.

[28] S. Federhen, "Type material in the NCBI taxonomy database," *Nucleic Acids Research*, vol. 43, Database issue, pp. D1086–D1098, 2015.

[29] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792–1797, 2004.

[30] C. E. Shannon, "The mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948, 623-656.

[31] M. Garcia-Boronat, C. M. Diez-Rivero, E. L. Reinherz, and P. A. Reche, "PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery," *Nucleic Acids Research*, vol. 36, Web Server issue, pp. W35–W41, 2008.

[32] P. A. Reche and E. L. Reinherz, "Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms," *Journal of Molecular Biology*, vol. 331, no. 3, pp. 623–641, 2003.

[33] J. J. Stewart, C. Y. Lee, S. Ibrahim et al., "A Shannon entropy analysis of immunoglobulin and T cell receptor," *Molecular Immunology*, vol. 34, pp. 1067–1082, 1997.

[34] P. A. Reche, J.-P. Glutting, and E. L. Reinherz, "Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles," *Immunogenetics*, vol. 56, pp. 405–419, 2004.

[35] P. A. Reche, J. P. Glutting, and E. L. Reinherz, "Prediction of MHC class I binding peptides using profile motifs," *Human Immunology*, vol. 63, no. 9, pp. 701–709, 2002.

[36] P. A. Reche and E. L. Reinherz, "Prediction of peptide-MHC binding using profiles," *Methods in Molecular Biology*, vol. 409, pp. 185–200, 2007.

[37] J. Greenbaum, J. Sidney, J. Chung, C. Brander, B. Peters, and A. Sette, "Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes," *Immunogenetics*, vol. 63, no. 6, pp. 325–335, 2011.

[38] P. Wang, J. Sidney, Y. Kim et al., "Peptide binding predictions for HLA DR, DP and DQ molecules," *BMC Bioinformatics*, vol. 11, p. 568, 2010.

[39] M. Molero-Abraham, E. M. Lafuente, D. R. Flower, and P. A. Reche, "Selection of conserved epitopes from hepatitis C virus for pan-populational stimulation of T-cell responses," *Clinical & Developmental Immunology*, vol. 2013, Article ID 601943, 10 pages, 2013.

[40] K. Cao, J. Hollenbach, X. Shi, W. Shi, M. Chopek, and M. A. Fernandez-Vina, "Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations," *Human Immunology*, vol. 62, no. 9, pp. 1009–1030, 2001.

[41] H. H. Bui, J. Sidney, K. Dinh, S. Southwood, M. J. Newman, and A. Sette, "Predicting population coverage of T-cell epitope-based diagnostics and vaccines," *BMC Bioinformatics*, vol. 7, p. 153, 2006.

[42] S. J. Hubbard and J. M. Thornton, *NACCESS, Computer Program*, Department of Biochemistry and Molecular Biology, University College London, London, England, UK, 1993.

[43] M. Magrane, "UniProt Knowledgebase: a hub of integrated protein data," *Database: The Journal of Biological Databases and Curation*, vol. 2011, article bar009, 2011.

[44] S. F. Altschul, T. L. Madden, A. A. Schaffer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.

[45] J. Peterson, S. Garges, M. Giovanni et al., "The NIH Human Microbiome Project," *Genome Research*, vol. 19, no. 12, pp. 2317–2323, 2009.

[46] L. H. Glimcher and C. J. Kara, "Sequences and factors: a guide to MHC class-II transcription," *Annual Review of Immunology*, vol. 10, pp. 13–49, 1992.

[47] L. J. Stern and J. M. Calvo-Calle, "HLA-DR: molecular insights and vaccine design," *Current Pharmaceutical Design*, vol. 15, no. 28, pp. 3249–3261, 2009.

[48] G. Szakonyi, M. G. Klein, J. P. Hannan et al., "Structure of the Epstein-Barr virus major envelope glycoprotein," *Nature Structural & Molecular Biology*, vol. 13, no. 11, pp. 996–1001, 2006.

[49] M. Backovic, R. Longnecker, and T. S. Jardetzky, "Structure of a trimeric variant of the Epstein-Barr virus glycoprotein B," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 8, pp. 2880–2885, 2009.

[50] E. Westhof, D. Altschuh, D. Moras et al., "Correlation between segmental mobility and the location of antigenic determinants in proteins," *Nature*, vol. 311, no. 5982, pp. 123–126, 1984.

[51] A. N. Kirschner, J. Sorem, R. Longnecker, and T. S. Jardetzky, "Structure of Epstein-Barr virus glycoprotein 42 suggests a mechanism for triggering receptor-activated virus entry," *Structure*, vol. 17, no. 2, pp. 223–233, 2009.

[52] K. Sathiyamoorthy, Y. X. Hu, B. S. Mohl, J. Chen, R. Longnecker, and T. S. Jardetzky, "Structural basis for Epstein-Barr virus host cell tropism mediated by gp42 and gHgL entry glycoproteins," *Nature Communications*, vol. 7, article 13557, 2016.

[53] C. Smith and R. Khanna, "The development of prophylactic and therapeutic EBV vaccines," *Current Topics in Microbiology and Immunology*, vol. 391, pp. 455–473, 2015.

[54] H. H. Balfour Jr., "Progress, prospects, and problems in Epstein-Barr virus vaccine development," *Current Opinion in Virology*, vol. 6, pp. 1–5, 2014.

[55] E. M. Sokal, K. Hoppenbrouwers, C. Vandermeulen et al., "Recombinant gp350 vaccine for infectious mononucleosis: a phase 2, randomized, double-blind, placebo-controlled trial to evaluate the safety, immunogenicity, and efficacy of an Epstein-Barr virus vaccine in healthy young adults," *The Journal of Infectious Diseases*, vol. 196, no. 12, pp. 1749–1753, 2007.

[56] R. Lin, D. Heeke, H. Liu et al., "Development of a robust, higher throughput green fluorescent protein (GFP)-based Epstein-Barr virus (EBV) micro-neutralization assay," *Journal of Virological Methods*, vol. 247, pp. 15–21, 2017.

[57] S. L. Elliott, A. Suhrbier, J. J. Miles et al., "Phase I trial of a $CD8^+$ T-cell peptide epitope-based vaccine for infectious

mononucleosis," *Journal of Virology*, vol. 82, no. 3, pp. 1448–1457, 2008.

[58] L. M. Hutt-Fletcher, "EBV glycoproteins: where are we now?," *Future Virology*, vol. 10, no. 10, pp. 1155–1162, 2015.

[59] V. Dasari, K. H. Bhatt, C. Smith, and R. Khanna, "Designing an effective vaccine to prevent Epstein-Barr virus-associated diseases: challenges and opportunities," *Expert Review of Vaccines*, vol. 16, no. 4, pp. 377–390, 2017.

[60] J. Ponomarenko and M. Van Regenmortel, "B-cell epitope prediction," in *Structural Bioinformatics*, pp. 849–879, John Wiley & Sons, Hoboken, NJ, USA, 2009.

[61] M. H. Van Regenmortel, "What is a B-cell epitope?," *Methods in Molecular Biology*, vol. 524, pp. 3–20, 2009.

[62] P. A. Reche and E. L. Reinherz, "Definition of MHC supertypes through clustering of MHC peptide-binding repertoires," *Methods in Molecular Biology*, vol. 409, pp. 163–173, 2007.

[63] P. Sun, H. Ju, Z. Liu et al., "Bioinformatics resources and tools for conformational B-cell epitope prediction," *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 943636, 11 pages, 2013.

[64] L. Potocnakova, M. Bhide, and L. B. Pulzova, "An introduction to B-cell epitope mapping and in silico epitope prediction," *Journal of Immunology Research*, vol. 2016, Article ID 6760830, 11 pages, 2016.

[65] M. J. Blythe and D. R. Flower, "Benchmarking B cell epitope prediction: underperformance of existing methods," *Protein Science*, vol. 14, no. 1, pp. 246–248, 2005.

[66] J. Gao and L. Kurgan, "Computational prediction of B cell epitopes from antigen sequences," *Methods in Molecular Biology*, vol. 1184, pp. 197–215, 2014.

[67] K. A. Young, A. P. Herbert, P. N. Barlow, V. M. Holers, and J. P. Hannan, "Molecular basis of the interaction between complement receptor type 2 (CR2/CD21) and Epstein-Barr virus glycoprotein gp350," *Journal of Virology*, vol. 82, no. 22, pp. 11217–11227, 2008.

[68] J. E. Tanner, M. Coincon, V. Leblond et al., "Peptides designed to spatially depict the Epstein-Barr virus major virion glycoprotein gp350 neutralization epitope elicit antibodies that block virus-neutralizing antibody 72A1 interaction with the native gp350 molecule," *Journal of Virology*, vol. 89, no. 9, pp. 4932–4941, 2015.

[69] V. Litwin, J. Gumperz, P. Parham, J. H. Phillips, and L. L. Lanier, "Specificity of HLA class I antigen recognition by human NK clones: evidence for clonal heterogeneity, protection by self and non-self alleles, and influence of the target cell type," *The Journal of Experimental Medicine*, vol. 178, no. 4, pp. 1321–1336, 1993.

[70] R. Arnon and T. Ben-Yedidia, "Old and new vaccine approaches," *International Immunopharmacology*, vol. 3, no. 8, pp. 1195–1204, 2003.

[71] J. L. Whitton, N. Sheng, M. B. Oldstone, and T. A. McKee, "A "string-of-beads" vaccine, comprising linked minigenes, confers protection from lethal-dose virus challenge," *Journal of Virology*, vol. 67, no. 1, pp. 348–352, 1993.

[72] S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, and G. P. Raghava, "In silico approach for predicting toxicity of peptides and proteins," *PLoS One*, vol. 8, no. 9, article e73957, 2013.

[73] E. M. Varypataki, N. Benne, J. Bouwstra, W. Jiskoot, and F. Ossendorp, "Efficient eradication of established tumors in mice with cationic liposome-based synthetic long-peptide vaccines," *Cancer Immunology Research*, vol. 5, no. 3, pp. 222–233, 2017.

[74] R. Kuai, L. J. Ochyl, K. S. Bahjat, A. Schwendeman, and J. J. Moon, "Designer vaccine nanodiscs for personalized cancer immunotherapy," *Nature Materials*, vol. 16, no. 4, pp. 489–496, 2017.

[75] M. Brazzoli, D. Magini, A. Bonci et al., "Induction of broad-based immunity and protective efficacy by self-amplifying mRNA vaccines encoding influenza virus hemagglutinin," *Journal of Virology*, vol. 90, no. 1, pp. 332–344, 2015.

[76] R. A. Schwendener, "Liposomes as vaccine delivery systems: a review of the recent advances," *Therapeutic Advances in Vaccines*, vol. 2, no. 6, pp. 159–182, 2014.

[77] D. Felnerova, J. F. Viret, R. Gluck, and C. Moser, "Liposomes and virosomes as delivery systems for antigens, nucleic acids and drugs," *Current Opinion in Biotechnology*, vol. 15, no. 6, pp. 518–529, 2004.

[78] Y. Perrie, D. Kirby, V. W. Bramwell, and A. R. Mohammed, "Recent developments in particulate-based vaccines," *Recent Patents on Drug Delivery & Formulation*, vol. 1, no. 2, pp. 117–129, 2007.

[79] C. Gujer, B. Chatterjee, V. Landtwing, A. Raykova, D. McHugh, and C. Munz, "Animal models of Epstein Barr virus infection," *Current Opinion in Virology*, vol. 13, pp. 6–10, 2015.

[80] C. Calcagno, R. Puzone, Y. E. Pearson et al., "Computer simulations of heterologous immunity: highlights of an interdisciplinary cooperation," *Autoimmunity*, vol. 44, no. 4, pp. 304–314, 2011.

[81] Proceedings of the IARC working group on the evaluation of carcinogenic risks to humans. Epstein-Barr virus and Kaposi's sarcoma herpesvirus/human herpesvirus 8. Lyon, France, 17-24 June 1997," *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*, vol. 70, pp. 1–492, 1997.

[82] A. W. Lee, W. Foo, O. Mang et al., "Changing epidemiology of nasopharyngeal carcinoma in Hong Kong over a 20-year period (1980-99): an encouraging reduction in both incidence and mortality," *International Journal of Cancer*, vol. 103, no. 5, pp. 680–685, 2003.

[83] W. Zhong, P. A. Reche, C. C. Lai, B. Reinhold, and E. L. Reinherz, "Genome-wide characterization of a viral cytotoxic T lymphocyte epitope repertoire," *The Journal of Biological Chemistry*, vol. 278, no. 46, pp. 45135–45144, 2003.

[84] E. M. Lafuente and P. A. Reche, "Prediction of MHC-peptide binding: a systematic and comprehensive overview," *Current Pharmaceutical Design*, vol. 15, no. 28, pp. 3209–3220, 2009.

Hindawi
Journal of Immunology Research
Volume 2017, Article ID 6412353, 14 pages
https://doi.org/10.1155/2017/6412353

Research Article

Vaccinomics Approach for Designing Potential Peptide Vaccine by Targeting *Shigella* spp. Serine Protease Autotransporter Subfamily Protein SigA

Arafat Rahman Oany,[1] Tahmina Pervin,[2] Mamun Mia,[1] Motaher Hossain,[3] Mohammad Shahnaij,[4] Shahin Mahmud,[1] and K. M. Kaderi Kibria[1]

[1]*Department of Biotechnology and Genetic Engineering, Faculty of Life Science, Mawlana Bhashani Science and Technology University, Tangail, Bangladesh*
[2]*Biotechnology and Genetic Engineering Discipline, Life Science School, Khulna University, Khulna, Bangladesh*
[3]*Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka, Bangladesh*
[4]*Enteric and Food Microbiology Laboratory, International Centre for Diarrhoeal Disease Research Bangladesh (icddr,b), Dhaka, Bangladesh*

Correspondence should be addressed to K. M. Kaderi Kibria; km_kibria@yahoo.com

Received 18 March 2017; Revised 28 June 2017; Accepted 24 July 2017; Published 7 September 2017

Academic Editor: Pedro A. Reche

Shigellosis, a bacillary dysentery, is closely associated with diarrhoea in human and causes infection of 165 million people worldwide per year. Casein-degrading serine protease autotransporter of enterobacteriaceae (SPATE) subfamily protein SigA, an outer membrane protein, exerts both cytopathic and enterotoxic effects especially cytopathic to human epithelial cell type-2 (HEp-2) and is shown to be highly immunogenic. In the present study, we have tried to impose the vaccinomics approach for designing a common peptide vaccine candidate against the immunogenic SigA of *Shigella* spp. At first, 44 SigA proteins from different variants of *S. flexneri*, *S. dysenteriae*, *S. boydii*, and *S. sonnei* were assessed to find the most antigenic protein. We retrieved 12 peptides based on the highest score for human leukocyte antigen (HLA) supertypes analysed by NetCTL. Initially, these peptides were assessed for the affinity with MHC class I and class II alleles, and four potential core epitopes VTARAGLGY, FHTVTVNTL, HTTWTLTGY, and IELAGTLTL were selected. From these, FHTVTVNTL and IELAGTLTL peptides were shown to have 100% conservancy. Finally, IELAGTLTL was shown to have the highest population coverage (83.86%) among the whole world population. In vivo study of the proposed epitope might contribute to the development of functional and unique widespread vaccine, which might be an operative alleyway to thwart dysentery from the world.

1. Background

Shigella is a Gram-negative, facultative anaerobic, nonmotile, nonspore forming, and rod-shaped true bacteria closely related to *Salmonella* and *Escherichia coli*. The resulting infection by this organism called shigellosis, also known as bacillary dysentery or Marlow syndrome, is most typically associated with diarrhoea and other gastrointestinal symptoms in humans. This pathogen is usually found in water that is contaminated with human feces within the setting of poor hygiene among kids of underneath 5 years old and is transmitted via the fecal-oral route. The infection will occur even if there is just a bodily function of only ten to one hundred microorganisms [1]. In each year, 165 million cases of *Shigella* infection are accounted worldwide, of that, 163 million take place in developing countries and ultimately result in millions of death [2]. Bangladesh has got the top rates of shigellosis according to the recent Global Enteric Multicenter Study (GEMS) in Asia. The output of this study has revealed that the *Shigella* is the third leading reason behind diarrhoea in children [3, 4].

Shigella species are usually classified into four serogroups: *S. dysenteriae* (12 serotypes), *S. flexneri* (6 serotypes), *S. boydii* (18 serotypes), and *S. sonnei* (one serotype) based on the

biochemical properties and group-specific O antigens within the outer portion of the semipermeable membrane. *S. dysenteriae*, *S. flexneri*, and *S. boydii* are physiologically similar in distinction to *S. sonnei*. Among them, *S. flexneri* is the most frequently isolated species globally and accounts for 60% of cases in the unindustrialized countries; *S. sonnei* causes 77% of cases in the industrialized countries [1].

The underlying therapeutic challenge to manage *Shigella* is its accrued resistance to most often used antibiotics like ampicillin, tetracycline, streptomycin, nalidixic acid, and sulfamethoxazole-trimethoprim [5]. Earlier, ciprofloxacin, a third-generation fluoroquinolone antibiotic, has been used effectively for the treatment of bacillary dysentery [6]. However, this antibiotic is no longer helpful for the treatment of bacillary dysentery in south Asian countries together with Bangladesh, because of the dissemination of fluoroquinolone-resistant *variety* and its equivalent clones across the countries [7, 8]. Hence, it is essential to find a sustainable approach like vaccinomics, which can elicit long-term and consistent immunological responses to fight against *Shigella*.

SigA is annotated in the she pathogenicity island of *Shigella*, encoding SigA protein which belongs to the serine protease autotransporter of enterobacteriaceae (SPATE) subgroup proteins. The autotransporter proteins of Gram-negative bacteria exhibit an N-terminal signal sequence, required for secretion across the inner membrane, and a C-terminal domain that forms an amphipathic β-barrel pore that allows passage of the functional domain across the outer membrane. This type of exporter proteins either remains attached to the cell surface or is released from the cell by proteolytic cleavage [9]. SigA is a multifunctional protein, able to degrade casein with cytotoxic and enterotoxic effects. Moreover, SigA is cytopathic for human epithelial type-2 (HEp-2) cells, causing morphological changes and loss of integrity of the cell monolayers, important for the pathologic process of *Shigella* [10]. The position of SigA in the chromosome made them less vulnerable to loss compare to the other virulence factors harbouring within the plasmid, and more exposure to the immune cells occurred by this secreted toxin [11]. Most importantly, this protein has been shown to be immunogenic following infection with *Shigella* [10]. The generalized modules of membrane antigen- (GMMA-) based outer membrane proteins including SigA were also shown to be highly immunogenic [12], which prompted us to target SigA as one of the best vaccine candidates and to design potential peptide vaccine covering all the *Shigella* spp. and most of the regions of the world.

Epitope-based immunizing agents are often an inexpensive choice to thwart enteric *Shigella* infection. The identification of specific epitopes derived from infectious pathogens has considerably advanced the event of epitope-based vaccines (EVs). Higher understanding of the molecular basis of substance recognition and human leukocyte antigen- (HLA-) binding motifs has resulted in the advancement of rationally designed vaccines that solely depends on algorithms predicting the peptide's binding to human HLA. The traditional process for the development of a vaccine is very complex compared to that of the epitope-based vaccine, and additionally, it is chemically stable, more specific, and free of any infectious or oncogenic potential hazard [13]. However, the invention of a wet laboratory-based candidate epitope is expensive and laborious that requires varied medicine experiments in the laboratory for the ultimate choice of epitopes. Hence, the interest for predicting epitopes by computational strategies, alternate in silico approaches among researchers, is growing bit by bit with reduced efforts.

Vaccinomics is the application of integrated knowledge from different disciplines including immunogenetics and immunogenomics to develop candidate next-generation vaccine and understand its immune response [14]. Currently, various vaccinomics databases are accessible for identification of distinctive B lymphocyte epitopes and HLA ligands with high sensitivity and specificity [15–17]. The vaccinomics approach has already proven its potency in identifying the conserved epitope in the case of human immunodeficiency virus [18], multiple sclerosis [19], tuberculosis [20], and malaria [21] with desired results. In our study, we have applied vaccinomics approaches for the screening of potentially conserved epitopes by targeting protein SigA.

2. Methods

The flow chart summarizing the protocols for the complete epitope prediction is illustrated in Figure 1.

2.1. Sequence Retrieval and Antigenic Protein Determination. The SigA protein sequences of different strains of *Shigella* species were retrieved from the NCBI GenBank [22] database and analysed in the VaxiJen v2.0 [23] server for the determination of the most potent antigenic protein. Additionally, the target protein was crosschecked against human pathogens and other similar pathogens to ensure the orthologous entry by using BLAST-P [24] and OrthoMCL [25] databases [26].

2.2. T-Cell Epitope Prediction and Affinity with MHC. The epitope prediction for the respective protein and their affinity score with MHC class I and class II allele was measured following previously used approach [27, 28]. Concisely, the NetCTL v1.2 server [29] was used for predicting potential cytotoxic T-lymphocyte (CTL) epitopes from the most antigenic protein. A combined algorithms including MHC-I binding, transporter of antigenic peptide (TAP) transport efficiency, and proteasomal C-terminal cleavage prediction were employed for the T-cell epitope prediction. The epitope with the highest score for 12 MHC class I supertypes was selected.

T Cell Epitope Prediction Tools from Immune Epitope Database and Analysis Resource (IEDB-AR) were used for the prediction of affinity with MHC class I [30] and MHC class II [31, 32]. The stabilized matrix method (SMM) was used to calculate the half-maximal inhibitory concentration (IC_{50}) of peptide binding to MHC class I with a preselected 9.0-mer epitope. The peptides were also assessed for HLA I binding affinity by the software, EPISOFT. For the analyses

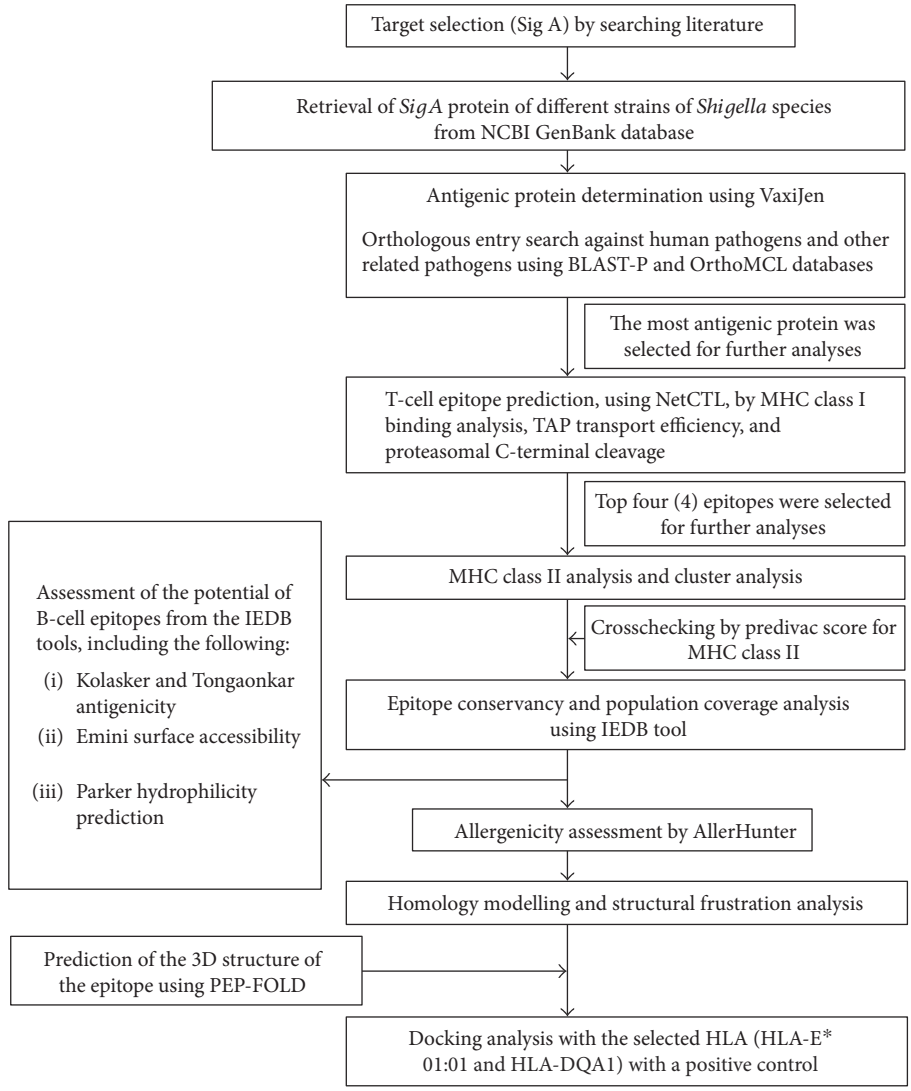


Figure 1: Flow diagram of the methodology.

of MHC class II binding, the IEDB-recommended method was used for the specific HLA-DP, HLA-DQ, and HLA-DR loci. Fifteen-mer epitopes were designed for MHC class II binding analysis considering the preselected 9-mer epitope and its conserved region in the *Shigella* strains. For the MHC class I and MHC class II alleles, the epitopes consisting $IC_{50} < 250$ nM and $IC_{50} < 100$ nM, respectively, were selected for further analysis. The MHC class II binding prediction tool PREDIVAC was also used to assess their affinity with HLA_DRB_1.

2.3. Cluster Analysis of the MHC Restricted Alleles. Furthermore, the MHCcluster v2.0 server [33] was used for the identification of cluster of MHC restricted allele with appropriate peptides to further strengthen our prediction. This is the additional crosscheck of the predicted MHC restricted allele analysis from the IEDB analysis resources. The output from this server is a static heat map and a graphical tree for describing the functional relationship between peptides and HLAs.

Table 1: T-cell epitopes of SigA protein predicted by the NetCTL server on the basis of the combined score. Here, epitopes for all the 12 different HLA supertypes have been presented.

Supertypes	Epitope	Start position	Combined score
A1	VTARAGLGY	645	2.8475
A2	SMYNTLWRV	279	1.4558
A3	KSYMSGNYK	432	1.5812
A24	NYAWVNGNI	219	1.6813
A26	HTTWTLTGY	411	2.0947
B7	APKGSNKEI	378	1.1981
B8	YMSGNYKAF	434	1.4879
B27	ARAGLGYQF	647	1.7245
B39	FHTVTVNTL	313	2.6855
B44	IELAGTLTL	178	1.9586
B58	IASFSQPDW	43	1.9421
B62	YMSGNYKAF	434	1.4814

Table 2: Epitopes for $CD8^+$ T-cell along with their interacting MHC class I alleles with affinity < 250 nM.

Epitope	Interacting MHC-I allele (IC_{50}) on the nM scale
VTARAGLGY	HLA-A∗29:02 (40.62), HLA-A∗01:01 (224.16), HLA-A∗30:02 (168.45), HLA-B∗15:02 (81.70), HLA-C∗03:03 (103.66)
FHTVTVNTL	HLA-B∗39:01 (9.02), HLA-B∗38:01 (62.10), HLA-B∗15:02 (21.84), HLA-C∗03:03 (5.48), HLA-C∗08:02 (211.91), HLA-C∗14:02 (92.18), HLA-C∗12:03 (78.92)
HTTWTLTGY	HLA-A∗26:01 (162.41), HLA-A∗25:01 (219.32), HLA-B∗15:02 (153.54), HLA-C∗14:02 (122.92), HLA-C∗12:03 (33.90), HLA-C∗03:03 (144.75)
IELAGTLTL	HLA-E∗01:01 (152.83), HLA-B∗40:01 (35.47), HLA-B∗15:02 (55.75), HLA-C∗03:03 (29.28), HLA-C∗12:03 (109.70)
IASFSQPDW	HLA-B∗58:01 (9.34), HLA-B∗57:01 (42.06), HLA-B∗53:01 (61.22), HLA-C∗03:03 (2.35), HLA-C∗08:02 (187.57), HLA-C∗15:02 (164.60), HLA-C∗12:03 (11.95), HLA-C∗14:02 (248.10)
ARAGLGYQF	HLA-C∗12:03, HLA-C∗03:03 (53.53), HLA-B∗15:02 (107.95), HLA-B∗27:05 (97.18)
NYAWVNGNI	HLA-A∗68:02 (188.19), HLA-C∗14:02 (21.26), HLA-C∗03:03 (220.60), HLA-C∗12:03 (60.56)
KSYMSGNYK	HLA-C∗15:02 (6.94), HLA-A∗30:01 (9.18), HLA-A∗11:01 (11.28), HLA-A∗03:01 (20.22), HLA-A∗31:01 (54.63), HLA-C∗14:02 (40.80), HLA-C∗03:03 (61.75), HLA-C∗12:03 (31.71)
YMSGNYKAF	HLA-B∗15:01 (61.09), HLA-B∗35:01 (124.62), HLA-C∗14:02 (17.24), HLA-B∗15:02 (61.69), HLA-C∗03:03 (13.14), HLA-C∗12:03 (151.43)
SMYNTLWRV	HLA-A∗02:01 (6.70), HLA-A∗02:06 (13.79), HLA-C∗14:02 (105.59), HLA-C∗12:03 (30.21)
APKGSNKEI	HLA-B∗07:02 (199.48), HLA-C∗12:03 (7.82), HLA-C∗03:03 (37.81)

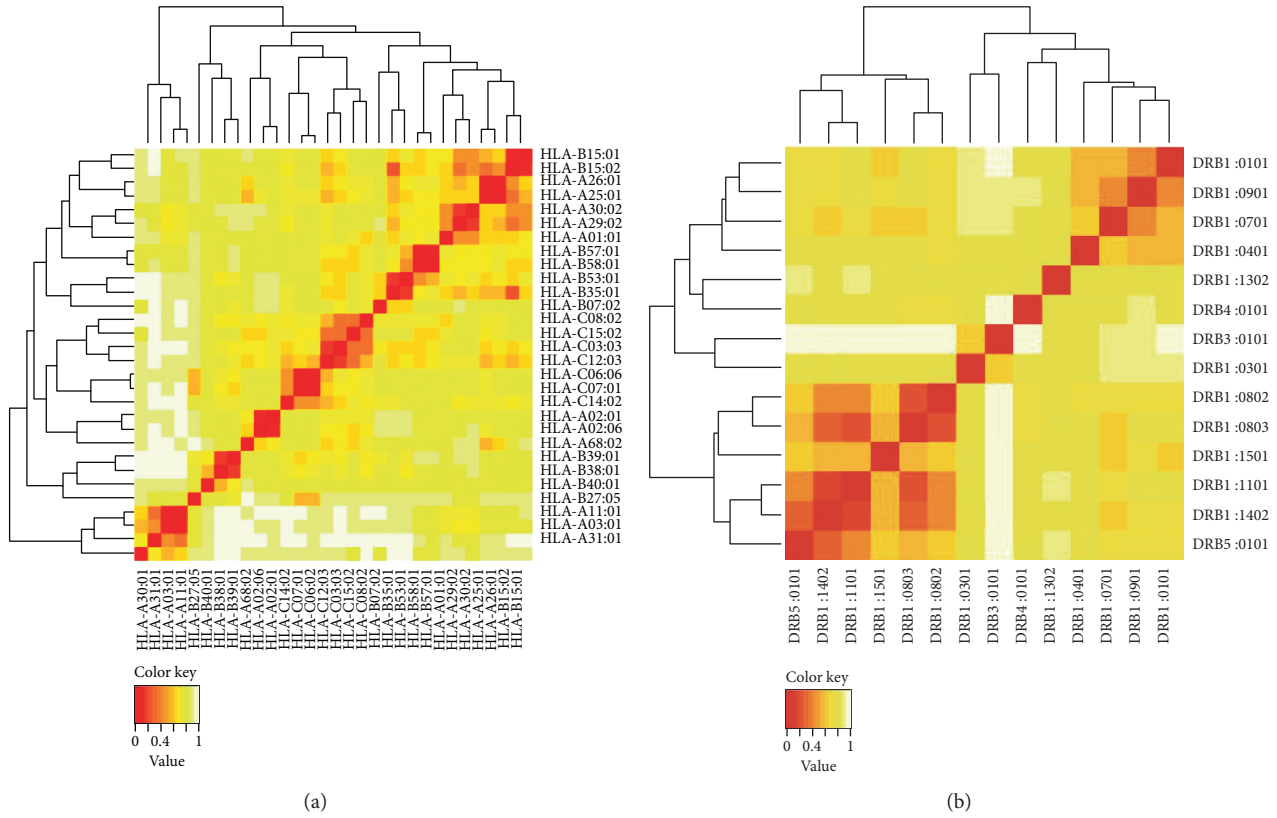


Figure 2: Cluster analysis of the HLA alleles for both MHC molecules through heat map representation. (a) Representing the cluster of the MHC-I. (b) Representing the cluster of MHC-II molecules. Epitopes are clustered on the basis of interaction with HLA and shown as red colour indicating strong interaction with appropriate annotation. Yellow zone indicates the weaker interaction. Here, all the available alleles are shown only.

2.4. Epitope Conservancy and Population Coverage Analyses. Epitope conservancy of the candidate epitopes was examined using a web-based epitope conservancy tool available in IEDB analysis resource [34]. The conservancy level of each potential epitope was calculated by considering identities in all SigA protein sequences of different strains retrieved from the database. Multiple sequence alignment (MSA) was employed to understand the positions of the

Table 3: The potential $CD4^+$ T-cell epitopes along with their interacting MHC class II alleles with affinity (IC50) < 100 nM and respective predivac scores.

Epitope	Interacting MHC-II allele (IC_{50}) on the nM scale	Number of alleles	Predivac score (binding core)
VTARAGLGYQFDLFA	HLA-DRB1∗04:01, HLA-DRB1∗09:01, HLA-DRB1∗04:05, HLA-DRB3∗01:01, HLA-DQA1∗05:01	5	68.76 (LGYQFDLFA)
NSGFHTVTVNTLDAT	HLA-DRB1∗01:01 (53), HLA-DRB1∗01:21 (17.87), HLA-DRB1∗01:17 (77.94), HLA-DRB1∗01:16 (98.93), HLA-DRB1∗01:13 (61.53), HLA-DRB1∗01:29 (83.08), HLA-DRB1∗01:24 (62.89), HLA-DRB1∗01:10 (28.18), HLA-DRB1∗01:11 (45.39), HLA-DRB1∗01:19 (39.99), HLA-DRB1∗01:12 (39.99), HLA-DRB1∗01:31 (39.99), HLA-DRB1∗01:32 (39.99), HLA-DRB1∗01:08 (39.99), HLA-DRB1∗01:05 (39.99), HLA-DRB1∗01:07 (39.99), HLA-DRB1∗01:27 (39.99), HLA-DRB1∗01:25 (39.99), HLA-DRB1∗01:22 (39.99), HLA-DRB1∗01:28 (39.99), HLA-DRB1∗01:14 (47.44), HLA-DRB1∗01:09 (37.12), HLA-DRB1∗01:15 (71.58), HLA-DRB1∗01:18 (44.25), HLA-DRB1∗01:06 (35.78), HLA-DRB1∗01:26 (38.4), HLA-DRB1∗01:20 (41.15), HLA-DRB1∗01:23 (48.69), HLA-DRB1∗01:04 (48.54), HLA-DRB1∗07:06 (65.25), HLA-DRB1∗07:05 (65.06), HLA-DRB5∗02:05 (46.61)	32	71.56 (FHTVTVNTL)
KAIELAGTLTLTGTP	HLA-DRB1∗01:01 (82), HLA-DRB1∗01:21 (23.85), HLA-DRB1∗01:17 (98.16), HLA-DRB1∗01:24 (58.9), HLA-DRB1∗01:10 (29.05), HLA-DRB1∗01:11 (49.8), HLA-DRB1∗01:19 (44.72), HLA-DRB1∗01:12 (44.72), HLA-DRB1∗01:31 (44.72), HLA-DRB1∗01:32 (44.72), HLA-DRB1∗01:08 (44.72), HLA-DRB1∗01:05 (44.72), HLA-DRB1∗01:07 (44.72), HLA-DRB1∗01:27 (44.72), HLA-DRB1∗01:25 (44.72), HLA-DRB1∗01:22 (44.72), HLA-DRB1∗01:28 (44.72), HLA-DRB1∗01:03 (91.98), HLA-DRB1∗01:14 (54.39), HLA-DRB1∗01:09 (31.87), HLA-DRB1∗01:15 (59.62), HLA-DRB1∗01:18 (40.48), HLA-DRB1∗01:06 (14.26), HLA-DRB1∗01:26 (18.43), HLA-DRB1∗01:20 (17.57), HLA-DRB1∗01:23 (24.27), HLA-DRB1∗01:04 (18.63), HLA-DRB1∗01:01 (82), HLA-DRB5∗02:05 (70.95)	29	71.70 (IELAGTLTL)
NNQIASFSQPDWESR		0	55.02 (FSQPDWESR)
VTARAGLGYQFDLFA		0	68.76 (LGYQFDLFA)
AQNYAWVNGNIKSDK	HLA-DRB5∗01:01 (62), HLA-DRB5∗02:04 (93.07), HLA-DRB5∗02:05 (24.55)	3	78.78 (YAWVNGNIK)
AAKSYMSGNYKAFLT	HLA-DRB1∗08:05 (65.65), HLA-DRB1∗12:03 (94.66), HLA-DRB1∗01:09 (57.66), HLA-DRB1∗01:10 (98.95), HLA-DRB1∗01:19 (99.23), HLA-DRB1∗01:12 (99.23), HLA-DRB1∗01:31 (99.23), HLA-DRB1∗01:32 (99.23), HLA-DRB1∗01:08 (99.23), HLA-DRB1∗01:05 (99.23), HLA-DRB1∗01:07 (99.23), HLA-DRB1∗01:27 (99.23), HLA-DRB1∗01:25 (99.23), HLA-DRB1∗01:22 (99.23), HLA-DRB1∗01:28 (99.23), HLA-DRB1∗01:21 (73.89), HLA-DRB1∗01:06 (52.24), HLA-DRB1∗01:23 (86.01), HLA-DRB1∗01:04 (87.14), HLA-DRB1∗01:26 (83.15), HLA-DRB1∗01:20 (85.84), HLA-DRB5∗02:05 (12.01), HLA-DRB5∗02:02 (67.2), HLA-DRB5∗02:04 (62.02)	24	60.30 (KSYMSGNYK)
SYMSGNYKAFLTEVN	HLA-DRB1∗01:21 (56.34), HLA-DRB1∗01:09 (89.52), HLA-DRB1∗01:10 (81.52), HLA-DRB1∗01:26 (98.98), HLA-DRB1∗01:06 (88.25), HLA-DRB1∗04:05 (38), HLA-DRB5∗02:05 (41.74)	7	69.98 (YKAFLTEVN)
ASMYNTLWRVNGQSA	HLA-DRB1∗08:05 (35.19), HLA-DRB1∗13:01 (49.45), HLA-DRB1∗12:05 (52.01), HLA-DRB1∗12:02 (60.35), HLA-DRB1∗12:03 (50.74), HLA-DRB1∗01:23 (62.08), HLA-DRB1∗01:26 (66.37), HLA-DRB1∗01:20 (74.19), HLA-DRB1∗01:06 (73.79), HLA-DRB5∗02:05 (44.37), HLA-DRB1∗11:01 (70)	11	82.45 (LWRVNGQSA)

epitopes within the sequences. As SPATE family is very much specific for the enterobacteria, specifically, *E. coli* and *Shigella*, we also include two *E. coli* sequences (gi|693049347| and gi|699401135|) along with those of four species of *Shigella* for MSA construction. The Jalview (http://www.jalview.org/) tool was used for this analysis. The conservancy of the selected peptides was also substantiated by the Protein Variability Software (PVS) [35]. Population coverage for the epitope was assessed by the IEDB population coverage calculation tool [36]. The combined score for MHC classes I and II was assessed for the analysis of the population coverage.

2.5. Homology Modelling and Structural Frustration Analysis. A homology model of the conserved region was obtained by MODELLER v9 [37], and the predicted model was assessed by the PROCHECK [38, 39] server. For the disorder prediction among the amino acid sequences, DISOPRED v3 [40]

Table 4: Conservancy analysis of all the epitopes with appropriate length.

Epitope	Conservancy	Length	Epitope	Conservancy	Length
VTARAGLGY	84.09%	9	**VTARAGLGY**QFDLFA	84.09%	15
SMYNTLWRV	100%	9	A**SMYNTLWRV**NGQSA	100%	15
KSYMSGNYK	97.73%	9	AA**KSYMSGNYK**AFLT	97.73%	15
NYAWVNGNI	100%	9	AQN**YAWVNGNI**KSDK	97.73%	15
HTTWTLTGY	97.73%	9	No prediction	Undetected	15
APKGSNKEI	100%	9	No prediction	Undetected	15
YMSGNYKAF	97.73%	9	S**YMSGNYKAF**LTEVN	97.73%	15
ARAGLGYQF	84.09%	9	VT**ARAGLGYQF**DLFA	84.09%	15
FHTVTVNTL	100%	9	NSG**FHTVTVNTL**DAT	100%	15
IELAGTLTL	100%	9	KA**IELAGTLTL**TGTP	100%	15
IASFSQPDW	97.73%	9	NNQ**IASFSQPDW**ESR	97.73%	15
YMSGNYKAF	97.73%	9	S**YMSGNYKAF**LTEVN	97.73%	15

was used. The protein frustratometer server [41] was employed for the detection of the stability and energy differences of the 3D structure of the protein.

2.6. Molecular Docking Analysis and HLA Allele Interaction. Docking studies were also performed using the best possible epitope following the strategy used in previous studies [27, 28]. AutoDock Vina [42] was used for the docking analysis. In our study, we have selected the HLA-E∗01:01 molecule as a candidate for MHC class I and the HLA-DQA1 as a candidate for MHC class II for docking analysis because they are the available hits in the Protein Data Bank (PDB) database. The PDB structure 2ESV, human cytomegalovirus complexes with T-cell receptors, VMAPRTLIL peptide, and 3PL6—structure of autoimmune TCR Hy.1B11 in complex with HLA-DQ1—were retrieved from the Research Collaboratory for Structural Bioinformatics (RCSB) protein database [43]. Then, the structures were simplified by using PyMOL (the PyMOL Molecular Graphics System, Version 1.5.0.4, Schrödinger, LLC) for the final docking purpose.

The PEP-FOLD server [44] was used for the conversion of the 3D structure of the epitope "IELAGTLTL" for MHC I and the epitope "KAIELAGTLTLTGTP" for the MHC II molecule in order to analyse the interaction with HLA alleles.

Finally, molecular docking was performed at the center of *X*: 77.8087, *Y*: −3.2264, and *Z*: −9.5769 and the dimensions (angstrom) of *X*: 31.4432, *Y*: 29.9517, and *Z*: 19.0455 for the MHC I molecules. For the MHC II molecules, docking was performed at the center of *X*: 38.5584, *Y*: 46.6132, and *Z*: −36.4392 and the dimensions (angstrom) of *X*: 34.8104, *Y*: 40.4401, and *Z*: 37.3366. Additionally, we have performed a control docking with the experimentally known peptide—MHC-bound complex. The PDB structure 2ESV, human cytomegalovirus complexes with T-cell receptors, and VMAPRTLIL peptide were used for this purpose. The gridline was used at the center of X: 77.3404, Y: −3.5159, and Z: −9.5829.

2.7. Allergenicity Investigation and B-Cell Epitope Prediction. The AllerHunter server [45] was used to predict the allergenicity of our proposed epitope for further securing the prediction, and the support vector machine (SVM) algorithm was used for the prediction within the server [46]. The predicted T-cell epitope (15-mer) was screened by IEDB-AR using a number of web-based tools for the suitability as the B-cell epitope [47–49].

3. Results

3.1. Analysis of the Retrieved Sequences and Their Antigenicity. A total of 44 SigA proteins from different variants of *S. flexneri*, *S. dysenteriae*, *S. boydii*, and *S. sonnei* were retrieved from the GenBank database (Table S1 in Supplementary Material available online at https://doi.org/10.1155/2017/6412353). Thereafter, analyses with the VaxiJen v2.0 server showed the protein with the accession number of gi|745767180| to have the highest antigenicity of 0.6699 (Table S1). This highly antigenic protein was further analysed to detect the highly immunogenic epitope. No significant entry was found in the orthologous entry search of our targeted protein.

3.2. T-Cell Epitope Identification. The NetCTLv1.2 server identified the T-cell epitopes, where the epitope prediction was confined to 12 MHC class I supertypes. Based on the combined score, the top twelve epitopes (Table 1) were listed for further analysis.

3.3. MHC Restriction and Cluster Analysis. IEDB analysis resource predicted both MHC class I and MHC class II restricted allele on the basis of the IC_{50} value. All the predicted epitopes in Table 1 were assessed for the MHC interaction analysis. Epitopes for the MHC class I alleles are presented in Table 2. The peptide IELAGTLTLT was predicted to have the highest number of MHC class I binding. This peptide was predicted to have the binding affinity with five MHC class I alleles including HLA-E∗01:01, HLA-B∗40:01, HLA-B∗15:02, HLA-C∗03:03, and HLA-C∗12:03. Furthermore, the interacted alleles were reassessed by cluster analysis and are shown in Figure 2(a), as a heat map, and in Figure S1A, as a dynamic tree. The peptides were reassessed by the EPISOPT software for the HLA I binding,

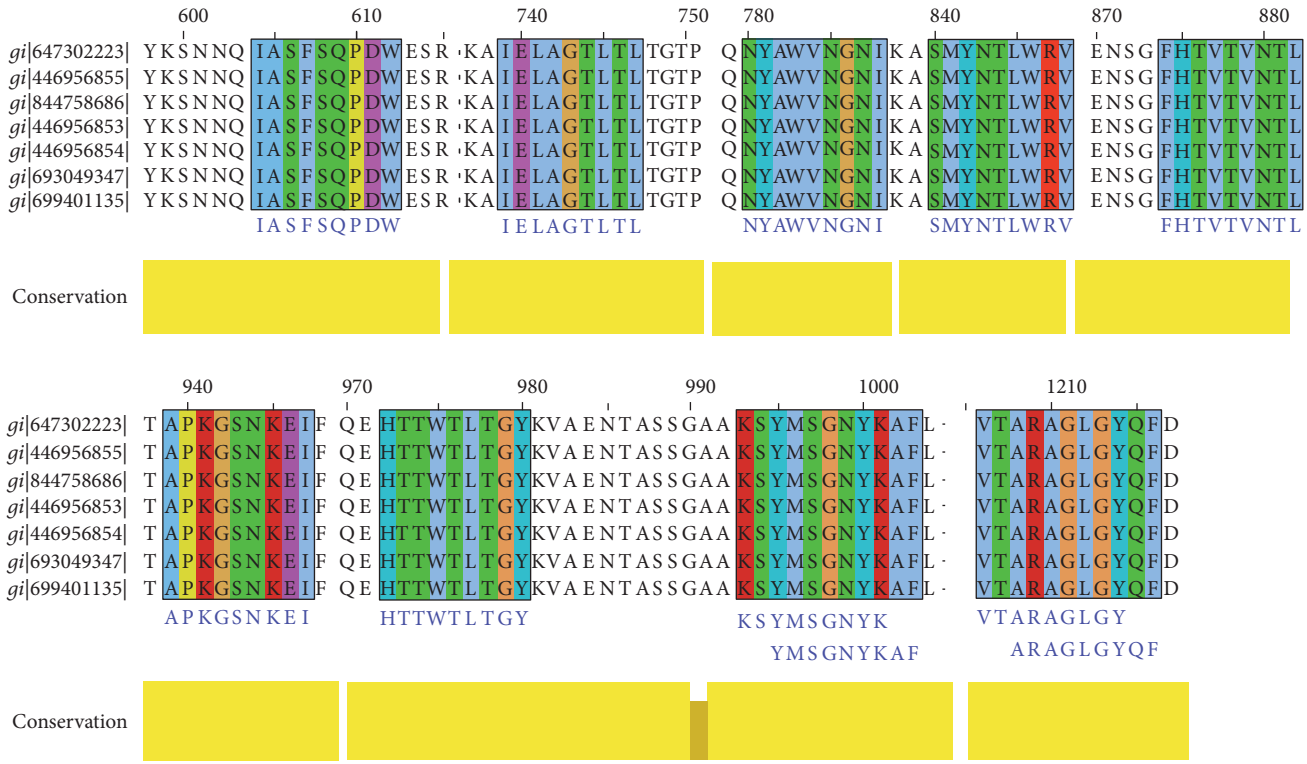


Figure 3: MSA-based location identification of the different epitopes within the SPATE proteins of *Shigella* and their homologue in *E. coli*. In this figure, gi|647302223|, gi|446956855|, gi|844758686|, and gi|446956853| represent the *S. flexneri*, *S. sonnei*, *S. boydii*, and *S. dysenteriae*, respectively. *E. coli* represented by gi|693049347| and gi|699401135|.

and IELAGTLTL was found to have affinity with six HLA I alleles (Table S3). From this analysis, we selected top four peptides VTARAGLGY, FHTVTVNTL, HTTWTLTGY, and IELAGTLTL depending on the affinity with most MHC class I.

Epitopes for the MHC class II alleles are presented in Table 3. Depending on the IC_{50} values as well as on the number of MHC class II alleles, three 15-mer peptide candidates were selected. The peptides NSG**FHTVTVNTL**DAT, KA**IELAGTLTL**TGTP, and AAKS**YMSGNYKAF**LT were predicted to have high affinity with MHC-II allele, which can interact with 32, 29, and 24 MHC class II alleles. The data has been validated by another software PREDIVAC. The predivac scores of the two core peptides FHTVTVNTL and IELAGTLTL have been shown to be promising for their binding to HLA_DRB_1 (Table 3). Accumulating both MHC class I allele- and MHC class II allele-based analyses, we showed FHTVTVNTL and IELAGTLTL peptides to have the best score to be a vaccine potential.

3.4. Conservancy Analysis and Position of the Epitopes. Conservancy of all the proposed epitopes was assessed by the IEDB conservancy analysis tool and is summarized in Table 4. FHTVTVNTL, IELAGTLTL, NYAWVNGNI, and SMYNTLWRV were shown to have 100% conserved regions across all the SigA proteins. The position of all the predicted epitopes is shown in a multiple sequence alignment of SigA proteins in Figure 3. Here, we used only our desired sequences for the proper annotation. So, from the most

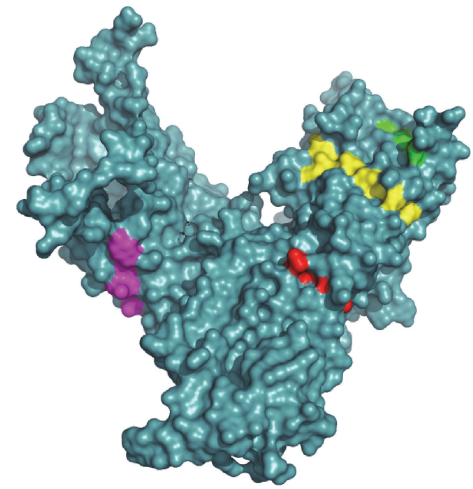
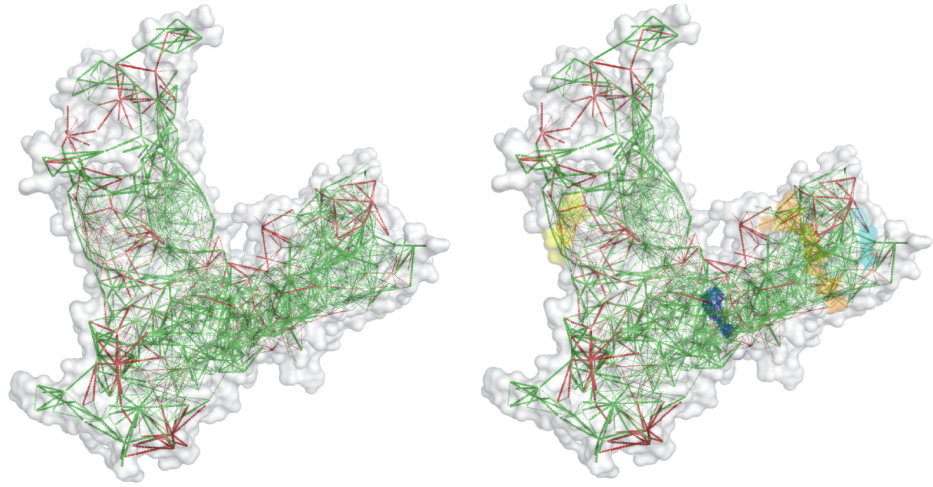


Figure 4: The three-dimensional model of SPATE subfamily protein SigA with the proposed epitopes VTARAGLGY (magenta), FHTVTVNTL (yellow), HTTWTLTGY (green), and IELAGTLTL (red). The superficial localities of the epitopes indicate their surface accessibility.

potential candidates, only two, that is, FHTVTVNTL and IELAGTLTL, were found to be fully conserved. The top four epitopes were shown within the protein in Figure 4. The conservancy of both of these peptides were crosschecked by PVS software, and it was found that they were located in the conserved region of the SigA protein (Figure S5).



(a) (b)

Figure 5: The configurational frustration index of the predicted model of the SigA. (a) This analysis detects the stability and energy differences of the 3D structure of the protein. Colours are in accordance with their frustration index. The red colour regions are highly frustrated and the green colour regions are not frustrated. The frustrated residues are able to change their identity and also displace the location in any favourable conditions. (b) The locations of our proposed epitopes are described by different colours. The epitopes HTTWTLTGY (cyan) and IELAGTLTL (blue) are well outside of the frustrated regions and securing their stability. On the other hand, the epitopes VTARAGLGY (yellow) and FHTVTVNTL (orange) are in the frustrated regions and unable to secure their stability.

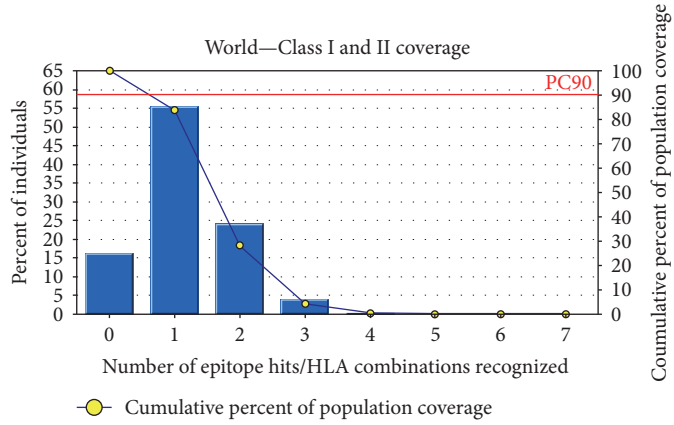


Figure 6: Population coverage analysis for the top predicted epitope based on the HLA interaction. Here, the whole world populations are assessed for the proposed epitope. The combined prediction for both of the MHC has been shown. Here, the number 1 bar for all the analyses represents out-predicted epitope. Notes: in the graphs, the line (-o-) represents the cumulative percentage of population coverage of the epitopes; the bars represent the population coverage for each epitope.

The epitopes are precisely positioned on the surface of the protein indicating that they would be accessible to the immune system, especially by B-cells.

3.5. Model Validation Structural Frustration Analysis. MODELLER modelled the three-dimensional structure of the targeted protein through the best multiple template-based modelling approach. The validation of the model was measured by the PROCHECK server through the Ramachandran plot and is depicted in Figure S2, where 88.8% amino acid residues were found within the favoured region. Furthermore, the predicted model was also assessed for the frustration analysis and is depicted in Figure 5. The DISOPRED server likewise assessed the disorder of the protein sequences in order to get an understanding about the disorder among the targeted sequences, which is shown in Figure S3.

3.6. Population Coverage Analysis. IEDB analysis resource predicted both MHC class I- and MHC class II-based coverage of the selected epitopes for the world population to assess the feasibility of being a potential vaccine candidate. The combined prediction was also assessed. The epitope "IELAGTLTLT" has the highest population coverage of 83.86% for the whole world population (shown graphically in Figure 6); however, another potential epitope "FHTVTVNTL" was shown to have 50.61% population coverage (Table S2).

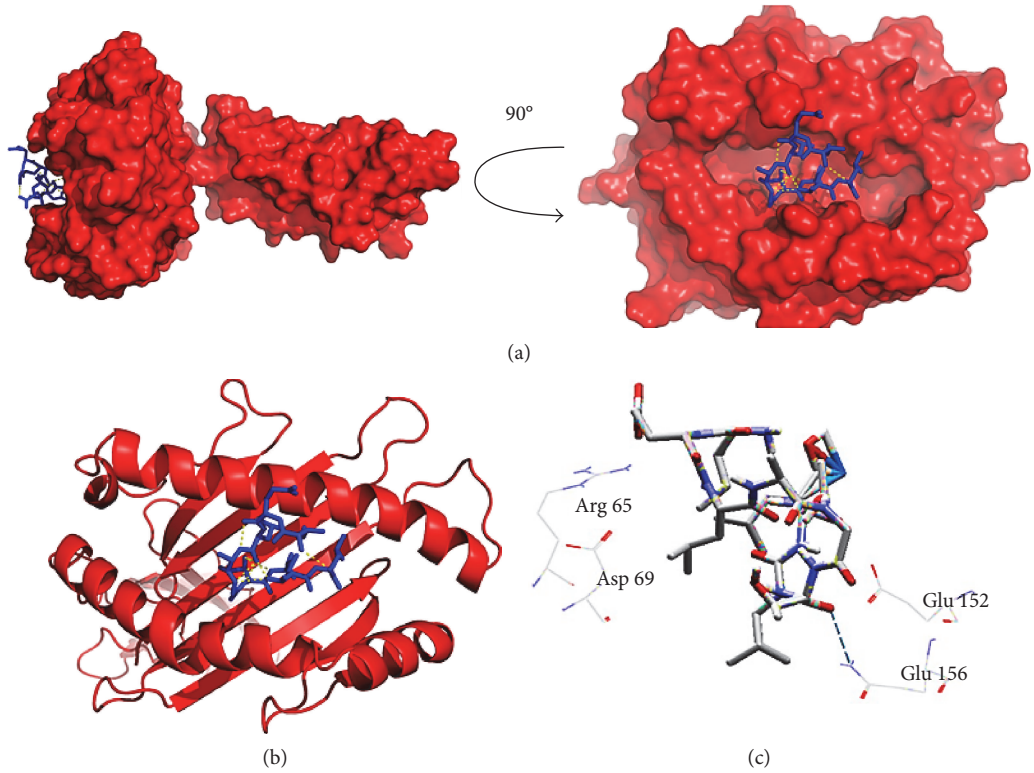


Figure 7: Docking analysis of the predicted epitope IELAGTLTL and HLA-E allele. (a) Representing the oriented view of the interaction and assuring the perfect binding. (b) Representing the cartoon view. (c) Embodying the interacted residues with the peptide.

3.7. Molecular Docking Analysis. The core epitope (IELAGTLTL) with 9.0 mer and its 15-mer extension (KAIELAGTLTLTGTP) were bound in the groove of the HLA-E∗01:01 and HLA-DQA1 with an energy of −7.8 and −9.7 kcal/mol, respectively. AutoDock Vina generated different poses of the docked peptide, and the best one was picked for the final calculation at an RMSD (root-mean-square deviation) value of 0.0. The docking interface was visualized with the PyMOL Molecular Graphics System. The 9.0-mer epitope interacted with Arg-61, Asn-62, and Glu-152 through steric interaction and formed hydrogen bonding with the Glu-156 amino acid residues. On the other hand, the 15-mer epitope interacted with Asp-55 through electrostatic interaction and Glu-66 through steric interaction and formed hydrogen bonding with the Gly-58, Arg-61, Asn-62, and Asn-82 amino acid residues. The docking output and the interacted residues are shown in Figures 7 and 8 with different orientations. Furthermore, the control docking energy was found to be −6.8 kcal/mol and is illustrated in Figure S4.

3.8. Allergenicity Analysis. The AllerHunter web server predicted the sequence-based allergenicity calculation very precisely. The allergenicity of the queried core epitope (IELAGTLTLT) was 0.05 (sensitivity = 98.40%, specificity = 27.4%), and the allergenicity of the 15-mer epitope (KAIELAGTLTLTGTP) was 0.05 (sensitivity = 98.4%, specificity = 27.0%).

3.9. B-Cell Epitope Prediction. B-cell epitope prediction was obtained for the peptide KAIELAGTLTLTGTP (15 mer) through the sequence-based approaches, and values are anticipated with different parameters, ranging from −0.6464 to 1.137. These values are the different propensity scores and predicted with a threshold ranging from −0.352 to 1.037 (Figure 9). The Kolaskar and Tongaonkar antigenicity scale was employed for evaluating the antigenic property of the peptide with a maximum of 1.072. The antigenic plot is showed in Figure 9(a). Peptide surface accessibility is another important benchmark to meet up the criteria of a potential B-cell epitope. Henceforth, Emini surface accessibility prediction was employed, with a maximum propensity score of 1.137 (Figure 9(b)). To reinforce our provision for the prediction of the epitope to elicit B-cell response, the Parker hydrophilicity prediction was also employed with a maximum score of 1.086 and is depicted in Figure 9(c).

4. Discussion

Enteric infections are the foremost cause of sickness and impermanence throughout the world, and only the *Shigella* infections resulted in over a million deaths annually [2]. The ever rising multidrug-resistant (MDR) strains of the *Shigella* bacteria area unit are another international concern for the researchers to search out a brand new resolution for preventing the deaths [50, 51]. Recently, there are several studies that focus on the development of the vaccine against *Shigella* and continue in the clinical trial. Most of them use attenuated and inactivated preparation of the bacteria for eliciting immune responses which has some potential escape risk [52–54]. In this study, we have

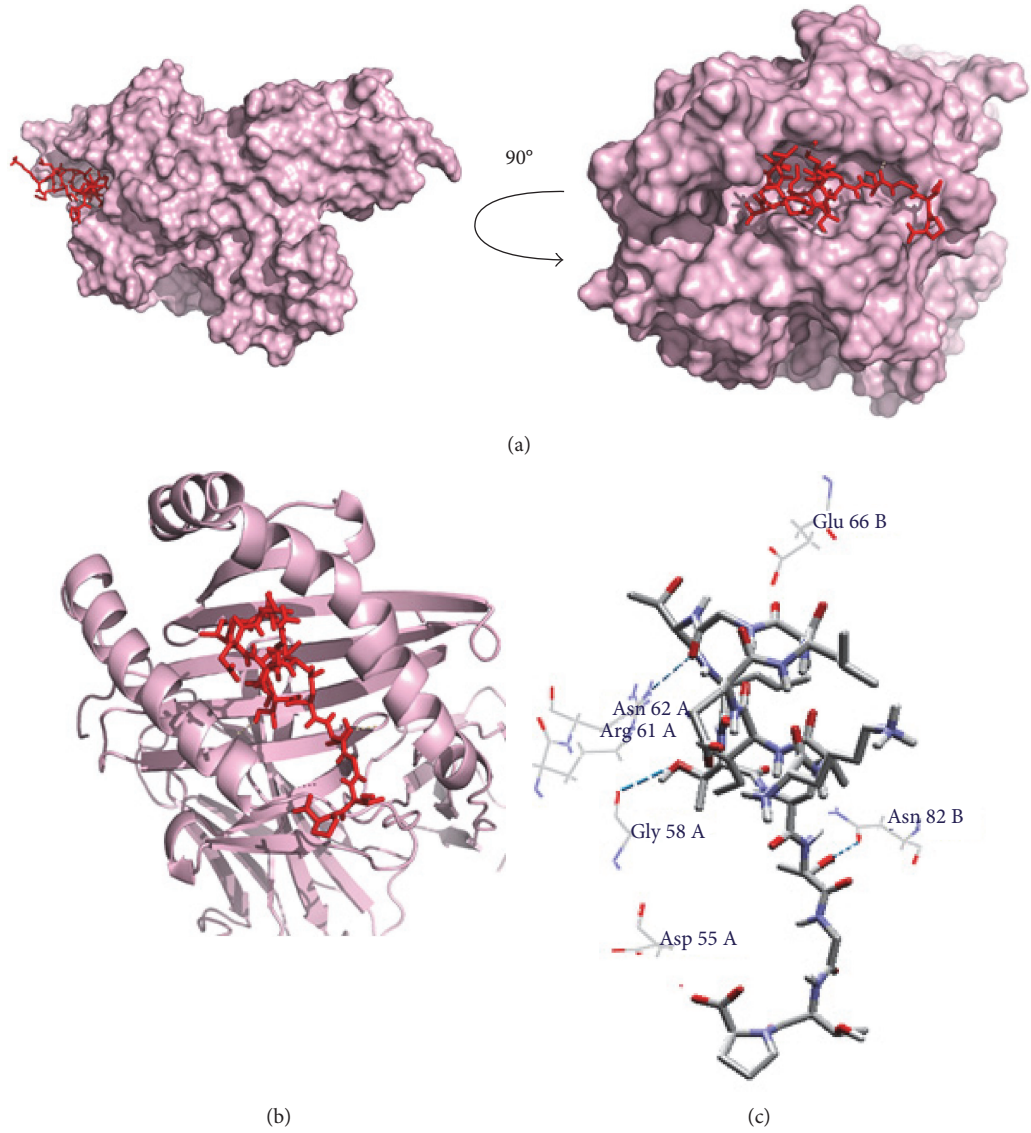


Figure 8: Docking analysis of the predicted epitope KAIELAGTLTLTGTP and HLA-DQA1 allele. (a) Representing the oriented view of the interaction and assuring the perfect binding. (b) Representing the cartoon view. (c) Embodying the interacted residues with the peptide.

tried to find out alternatives to treat this global burden through vaccinomics approaches and targeting the immunogenic and toxic protein SigA. The sequences of different strains of *Shigella* showed that there is a little island of conserved sequence throughout the species [55], and we have focused on that target for designing the vaccine candidate. The orthologous entry search of our targeted protein revealed no significant similarity with human pathogens and other closely related pathogens. These results further strengthen our prediction through confirming no cross immunity.

In recent time, most of the vaccines are grounded on B-cell immunity; vaccines based on a T-cell epitope have been invigorated lately. This is often as a result of body substance response from memory B-cells which may be overawed basically by matter drift as time goes on, whereas cell-mediated immunity repeatedly delivers long-run immunity [56, 57]. As a consequence, a T-lymphocyte epitope elicits a robust and distinctive immune response through the cytotoxic lymphocyte- (CTL-) mediated pathway and impedes the spreading of the infectious agents by the CTL through recognizing and killing the infected cells or by secreting specific cytokines [58].

The epitopes VTARAGLGY, FHTVTVNTL, HTTW TLTGY, and IELAGTLTL are primarily selected for the designing of vaccine from the initial analysis depending on the affinity with MHC class I and additionally confirmed their presence along with those of the ancestral homologue in *E. coli* (Figure 2). Finally, through substantiation with different parameters, the core epitopes IELAGTLTL and FHTVTVNTL (in 15.0-mer form, KAIELAGTLTLTGTP and NSGFHTVTVNTLDAT, resp.) were found to be the most potential and highly interacting HLA candidates for MHC class II molecule. Furthermore, we have used

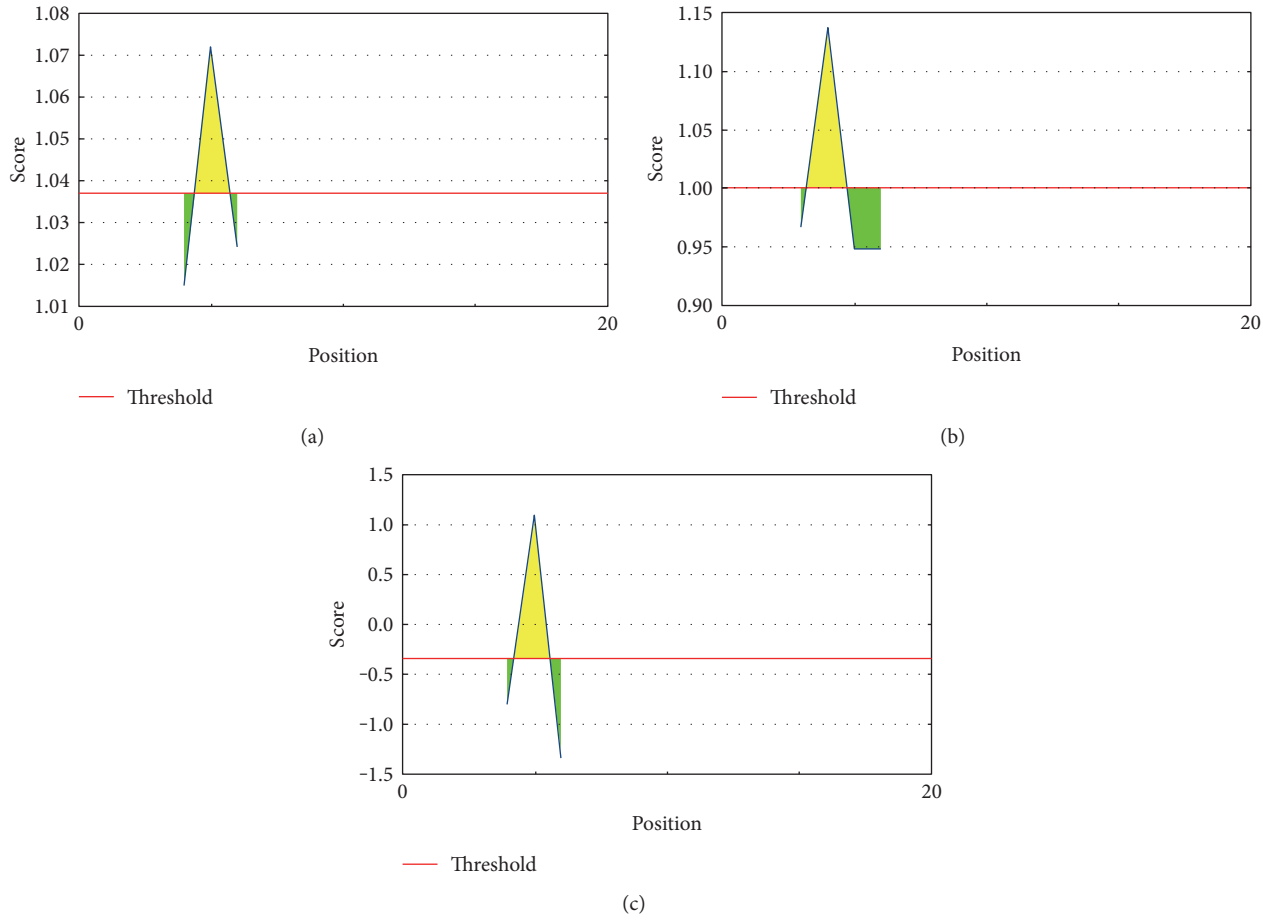


Figure 9: B-cell epitope prediction. (a) Kolaskar and Tongaonkar antigenicity prediction of the proposed epitope with a threshold value of 1.037. (b) Emini surface accessibility prediction of the proposed epitope, with a threshold value of 1.0. (c) Parker hydrophilicity prediction of the epitope, with a threshold of −0.352. Notes: the *x*-axis and *y*-axis represent the sequence position and antigenic propensity, respectively. The regions above the threshold are antigenic (desired), shown in yellow.

pSORTb to predict the subcellular localization of SigA and found that there is a score of 5.87 for localization in the outer membrane and another score of 4.13 for extracellular localization. The result was quite similar with that for the localization of other SPATE proteins in the bacterial cell surface as well as in secreted forms.

The three-dimensional model built through MODELLER and validated by the Ramachandran plot with an acceptable range resulted in the display of the perfect position of the epitope on the surface of the structure. As the epitope was found on the surface (Figure 4) of the model, it would increase the possibility to interact with the immune system earlier. Furthermore, the analysis from the DISOPRED and frustration analysis servers strengthen our prediction, though there are no disorder and energy frustration in the epitope region of the sequences and model, respectively (Figure 5 and Figure S3).

To get the acceptability, vaccine candidates must have wider population coverage. This is very much important before designing. In our analysis, we have found that our proposed epitope IELAGTLTL had combined population coverage of 83.86%, whereas the other most potential candidate FHTVTVNTL had combined population coverage of 50.61%. This output revealed that the proposed epitopes would have wider coverage *in vitro*.

Molecular docking upkeeps the prediction with a higher docking score and the perfectly oriented interactions between the both MHC and the predicted 9.0-mer and 15-mer epitopes. Additionally, comparative analysis with the experimentally known peptide—MHC complex—has also revealed the precision of our prediction through the similar binding energy and interacted residues. Another significant finding is the conservancy result. Through analysis of the whole retrieved sequences, it was found that our predicted epitopes have a 100% conservancy and hopefully they would be potential candidates for treating all of the *Shigella* spp. Our proposed epitopes are nonallergenic in nature according to the FAO/WHO allergenicity evaluation scheme.

Finally, the core epitope "IELAGTLTL" was also found to be more potential B-cell epitope candidates that were proposed through the sequence-based approaches including the Kolaskar and Tongaonkar antigenicity scale, Emini surface accessibility prediction, and Parker hydrophilicity prediction. From the overhead analysis, we envisage that our suggested epitope would also elicit an immune response *in vitro*.

5. Conclusion

The improved knowledge about antigen recognition at molecular level led us to the development of rationally designed peptide vaccines. The idea of peptide vaccines is based on detecting and chemical synthesis of immunodominant B-cell and T-cell epitopes capable of evoking specific immune responses. In this study, we used different computational tools to identify potential epitope targets against *Shigella* which will help to decrease the cost and time of wet lab experiments more successfully. Our bioinformatic analyses speculate that the selected part of the outer membrane and highly immunogenic protein, SigA, is a potential candidate for a peptide vaccine. It might also contribute to the reduction in the SigA-mediated pathogenicity to the host. However, further wet lab validation is necessary to confirm the efficiency of our identified peptide sequence as an epitope vaccine against *Shigella*.

Abbreviations

SPATE: Serine protease autotransporter of enterobacteria
HEp-2: Human epithelial cell type-2
MHC-I: Major histocompatibility complex class I
MHC-II: Major histocompatibility complex class II
MSA: Multiple sequence alignment
GMMA: Generalized modules of membrane antigens
CTL: Cytotoxic T-lymphocyte
TAP: Transporter of antigenic peptide
SMM: Stabilized matrix method
HLA: Human leukocyte antigen.

Additional Points

Availability of Data and Materials. Information about the data and their availability is intricately described in Methods.

Ethical Approval

As samples from human or animals had not been used in this study, ethical clearance is not applicable.

Consent

Patient consent is not applicable.

Conflicts of Interest

No potential competing interest was reported by the authors.

Authors' Contributions

Arafat Rahman Oany conceived, designed, and guided the study; drafted the manuscript; and analysed the data. Tahmina Pervin, Mamun Mia, and Motaher Hossain carried out the molecular genetic studies, participated in the sequence alignment, and drafted the manuscript. Mohammad Shahnaij helped in the design of the study. Shahin Mahmud helped in drafting the manuscript. K. M. Kaderi Kibria participated in the design and coordination, performed critical revision, and helped in drafting the manuscript. All authors read and approved the final manuscript.

References

[1] T. L. Hale and G. T. Keusch, "Shigella," in *Medical Microbiology*, S. Baron, Ed., University of Texas Medical Branch at Galveston, Galveston (TX), USA, 1996, Chapter 22.

[2] K. L. Kotloff, J. P. Winickoff, B. Ivanoff et al., "Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies," *Bulletin of the World Health Organization*, vol. 77, no. 8, p. 651, 1999.

[3] L. von Seidlein, D. R. Kim, M. Ali et al., "A multicentre study of *Shigella* diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology," *PLoS Medicine*, vol. 3, no. 9, article e353, 2006.

[4] K. L. Kotloff, J. P. Nataro, W. C. Blackwelder et al., "Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study," *Lancet*, vol. 382, no. 9888, p. 209, 2003.

[5] S. Dutta, K. Rajendran, S. Roy et al., "Shifting serotypes, plasmid profile analysis and antimicrobial resistance pattern of shigellae strains isolated from Kolkata, India during 1995-2000," *Epidemiology and Infection*, vol. 129, no. 2, p. 235, 2002.

[6] M. A. Salam, U. Dhar, W. A. Khan, and M. L. Bennish, "Randomised comparison of ciprofloxacin suspension and pivmecillinam for childhood shigellosis," *Lancet*, vol. 352, no. 9127, p. 522, 1998.

[7] K. A. Talukder, B. K. Khajanchi, M. A. Islam et al., "Genetic relatedness of ciprofloxacin-resistant *Shigella dysenteriae* type 1 strains isolated in south Asia," *The Journal of Antimicrobial Chemotherapy*, vol. 54, no. 4, p. 730, 2004.

[8] K. A. Talukder, B. K. Khajanchi, M. A. Islam et al., "Fluoroquinolone resistance linked to both gyrA and parC mutations in the quinolone resistance-determining region of *Shigella dysenteriae* type 1," *Current Microbiology*, vol. 52, no. 2, p. 108, 2006.

[9] I. R. Henderson, F. Navarro-Garcia, and J. P. Nataro, "The great escape: structure and function of the autotransporter proteins," *Trends in Microbiology*, vol. 6, no. 9, p. 370, 1998.

[10] K. Al-Hasani, F. Navarro-Garcia, J. Huerta, H. Sakellaris, and B. Adler, "The immunogenic SigA enterotoxin of *Shigella flexneri* 2a binds to HEp-2 cells and induces fodrin redistribution in intoxicated epithelial cells," *PLoS One*, vol. 4, no. 12, article e8223, 2009.

[11] K. Al-Hasani, I. R. Henderson, H. Sakellaris et al., "The *sigA* gene which is borne on the *she*pathogenicity island of *Shigella flexneri* 2a encodes an exported cytopathic protease involved in intestinal fluid accumulation," *Infection and Immunity*, vol. 68, no. 5, p. 2457, 2000.

[12] F. Berlanda Scorza, A. M. Colucci, L. Maggiore et al., "High yield production process for *Shigella* outer membrane particles," *PLoS One*, vol. 7, no. 6, article e35616, 2012.

[13] A. Sette, M. Newman, B. Livingston et al., "Optimizing vaccine design for cellular processing, MHC binding and TCR recognition," *Tissue Antigens*, vol. 59, no. 6, p. 443, 2002.

[14] G. A. Poland, I. G. Ovsyannikova, and R. M. Jacobson, "Application of pharmacogenomics to vaccines," *Pharmacogenomics*, vol. 10, no. 5, p. 837, 2009.

[15] S. P. Singh and B. N. Mishra, "Major histocompatibility complex linked databases and prediction tools for designing vaccines," *Human Immunology*, vol. 77, no. 3, p. 295, 2016.

[16] V. Brusic, G. Rudy, and L. C. Harrison, "MHCPEP, a database of MHC-binding peptides: update 1997," *Nucleic Acids Research*, vol. 26, no. 1, p. 368, 1998.

[17] H. Rammensee, J. Bachmann, N. P. Emmerich, O. A. Bachor, and S. Stevanovic, "SYFPEITHI: database for MHC ligands and peptide motifs," *Immunogenetics*, vol. 50, no. 3-4, p. 213, 1999.

[18] C. C. Wilson, D. McKinney, M. Anders et al., "Development of a DNA vaccine designed to induce cytotoxic T lymphocyte responses to multiple conserved epitopes in HIV-1," *Journal of Immunology*, vol. 171, no. 10, p. 5611, 2003.

[19] D. N. Bourdette, E. Edmonds, C. Smith et al., "A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis," *Multiple Sclerosis*, vol. 11, no. 5, p. 552, 2005.

[20] H. L. Robinson and R. R. Amara, "T cell vaccines for microbial infections," *Nature Medicine*, vol. 11, article S25, Supplement 4, 2005.

[21] J. A. Lopez, C. Weilenman, R. Audran et al., "A synthetic malaria vaccine elicits a potent $CD8^+$ and $CD4^+$ T lymphocyte immune response in humans. Implications for vaccination strategies," *European Journal of Immunology*, vol. 31, no. 7, p. 1989, 2001.

[22] D. A. Benson, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers, "GenBank," *Nucleic Acids Research*, vol. 43, article D30, Database issue, 2008.

[23] I. A. Doytchinova and D. R. Flower, "VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines," *BMC Bioinformatics*, vol. 8, p. 4, 2007.

[24] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, p. 403, 1990.

[25] F. Chen, A. J. Mackey, C. J. Stoeckert Jr., and D. S. Roos, "OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups," *Nucleic Acids Research*, vol. 34, article D363, Database issue, 2006.

[26] S. P. Singh, K. Roopendra, and B. N. Mishra, "Genome-wide prediction of vaccine candidates for *Leishmania major*: an integrated approach," *Journal of Tropical Medicine*, vol. 2015, Article ID 709216, 14 pages, 2015.

[27] A. R. Oany, A. A. Emran, and T. P. Jyoti, "Design of an epitope-based peptide vaccine against spike protein of human coronavirus: an in silico approach," *Drug Design, Development and Therapy*, vol. 8, p. 1139, 2014.

[28] A. R. Oany, T. Sharmin, A. S. Chowdhury, T. P. Jyoti, and M. A. Hasan, "Highly conserved regions in Ebola virus RNA dependent RNA polymerase may be act as a universal novel peptide vaccine target: a computational approach," *In Silico Pharmacol*, vol. 3, no. 1, p. 7, 2015.

[29] M. V. Larsen, C. Lundegaard, K. Lamberth, S. Buus, O. Lund, and M. Nielsen, "Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction," *BMC Bioinformatics*, vol. 8, p. 424, 2007.

[30] S. Buus, S. L. Lauemoller, P. Worning et al., "Sensitive quantitative predictions of peptide-MHC binding by a 'Query by Committee' artificial neural network approach," *Tissue Antigens*, vol. 62, no. 5, p. 378, 2003.

[31] P. Wang, J. Sidney, C. Dow, B. Mothe, A. Sette, and B. Peters, "A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach," *PLoS Computational Biology*, vol. 4, no. 4, article e1000048, 2008.

[32] P. Wang, J. Sidney, Y. Kim et al., "Peptide binding predictions for HLA DR, DP and DQ molecules," *BMC Bioinformatics*, vol. 11, p. 568, 2010.

[33] M. Thomsen, C. Lundegaard, S. Buus, O. Lund, and M. Nielsen, "MHCcluster, a method for functional clustering of MHC molecules," *Immunogenetics*, vol. 65, no. 9, p. 655, 2013.

[34] H. H. Bui, J. Sidney, W. Li, N. Fusseder, and A. Sette, "Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines," *BMC Bioinformatics*, vol. 8, p. 361, 2007.

[35] M. Garcia-Boronat, C. M. Diez-Rivero, E. L. Reinherz, and P. A. Reche, "PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery," *Nucleic Acids Research*, vol. 36, article W35, Web Server issue, 2008.

[36] H. H. Bui, J. Sidney, K. Dinh, S. Southwood, M. J. Newman, and A. Sette, "Predicting population coverage of T-cell epitope-based diagnostics and vaccines," *BMC Bioinformatics*, vol. 7, p. 153, 2006.

[37] A. Sali, L. Potterton, F. Yuan, H. van Vlijmen, and M. Karplus, "Evaluation of comparative protein modeling by MODELLER," *Proteins*, vol. 23, no. 3, p. 318, 1995.

[38] R. A. Laskowski, J. A. Rullmannn, M. W. MacArthur, R. Kaptein, and J. M. Thornton, "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR," *Journal of Biomolecular NMR*, vol. 8, no. 4, p. 477, 1996.

[39] K. Arnold, L. Bordoli, J. Kopp, and T. Schwede, "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling," *Bioinformatics*, vol. 22, no. 2, p. 195, 2006.

[40] J. J. Ward, L. J. McGuffin, K. Bryson, B. F. Buxton, and D. T. Jones, "The DISOPRED server for the prediction of protein disorder," *Bioinformatics*, vol. 20, no. 13, p. 2138, 2004.

[41] M. Jenik, R. G. Parra, L. G. Radusky, A. Turjanski, P. G. Wolynes, and D. U. Ferreiro, "Protein frustratometer: a tool to localize energetic frustration in protein molecules," *Nucleic Acids Research*, vol. 40, article W348, Web Server issue, 2012.

[42] O. Trott and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *Journal of Computational Chemistry*, vol. 31, no. 2, p. 455, 2010.

[43] H. M. Berman, J. Westbrook, Z. Feng et al., "The Protein Data Bank," *Nucleic Acids Research*, vol. 28, no. 1, p. 235, 2000.

[44] P. Thevenet, Y. Shen, J. Maupetit, F. Guyon, P. Derreumaux, and P. Tuffery, "PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides," *Nucleic Acids Research*, vol. 40, article W288, Web Server issue, 2012.

[45] H. C. Muh, J. C. Tong, and M. T. Tammi, "AllerHunter: a SVM-pairwise system for assessment of allergenicity and allergic cross-reactivity in proteins," *PLoS One*, vol. 4, no. 6, article e5861, 2009.

[46] L. Liao and W. S. Noble, "Combining pairwise sequence similarity and support vector machines for detecting remote protein evolutionary and structural relationships," *Journal of Computational Biology*, vol. 10, no. 6, p. 857, 2003.

[47] A. S. Kolaskar and P. C. Tongaonkar, "A semi-empirical method for prediction of antigenic determinants on protein antigens," *FEBS Letters*, vol. 276, no. 1-2, p. 172, 1990.

[48] E. A. Emini, J. V. Hughes, D. S. Perlow, and J. Boger, "Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide," *Journal of Virology*, vol. 55, no. 3, p. 836, 1985.

[49] J. M. Parker, D. Guo, and R. S. Hodges, "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*, vol. 25, no. 19, p. 5425, 1986.

[50] D. A. Rowe-Magnus and D. Mazel, "The role of integrons in antibiotic resistance gene capture," *International Journal of Medical Microbiology*, vol. 292, no. 2, p. 115, 2002.

[51] K. Goh, D. Chua, B. Beck, M. L. McKee, and A. A. Bhagwat, "Arginine-dependent acid-resistance pathway in *Shigella boydii*," *Archives of Microbiology*, vol. 193, no. 3, p. 179, 2011.

[52] R. Walker, "New possibilities for the development of a combined vaccine against ETEC and *Shigella*," *BMJ Global Health*, vol. 1, no. 2, article A11, Supplement 2, 2017.

[53] M. S. Riddle, R. W. Kaminski, C. Di Paolo et al., "Safety and immunogenicity of a candidate bioconjugate vaccine against *Shigella flexneri* 2a administered to healthy adults: a single-blind, randomized phase I study," *Clinical and Vaccine Immunology*, vol. 23, no. 12, p. 908, 2016.

[54] M. S. Riddle, R. W. Kaminski, C. Williams et al., "Safety and immunogenicity of an intranasal *Shigella flexneri* 2a invaplex 50 vaccine," *Vaccine*, vol. 29, no. 40, p. 7009, 2011.

[55] Q. Jin, Z. Yuan, J. Xu et al., "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157," *Nucleic Acids Research*, vol. 30, no. 20, p. 4432, 2002.

[56] R. Bacchetta, S. Gregori, and M. G. Roncarolo, "$CD4^+$ regulatory T cells: mechanisms of induction and effector function," *Autoimmunity Reviews*, vol. 4, no. 8, p. 491, 2005.

[57] J. U. Igietseme, F. O. Eko, Q. He, and C. M. Black, "Antibody regulation of T-cell immunity: implications for vaccine strategies against intracellular pathogens," *Expert Review of Vaccines*, vol. 3, no. 1, p. 23, 2004.

[58] B. Shrestha and M. S. Diamond, "Role of $CD8^+$ T cells in control of West Nile virus infection," *Journal of Virology*, vol. 78, no. 15, p. 8312, 2004.

Hindawi
Journal of Immunology Research
Volume 2017, Article ID 7407136, 16 pages
https://doi.org/10.1155/2017/7407136

Research Article

Fusion to Flaviviral Leader Peptide Targets HIV-1 Reverse Transcriptase for Secretion and Reduces Its Enzymatic Activity and Ability to Induce Oxidative Stress but Has No Major Effects on Its Immunogenic Performance in DNA-Immunized Mice

Anastasia Latanova,[1,2,3] Stefan Petkov,[3] Yulia Kuzmenko,[1] Athina Kilpeläinen,[3] Alexander Ivanov,[1] Olga Smirnova,[1] Olga Krotova,[1,2] Sergey Korolev,[4] Jorma Hinkula,[5] Vadim Karpov,[1] Maria Isaguliants,[2,6,7] and Elizaveta Starodubova[1,3,7]

[1]*Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*
[2]*Gamaleja Research Center of Epidemiology and Microbiology, Moscow, Russia*
[3]*Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden*
[4]*Chemistry Department, Belozersky Research Institute of Physico-Chemical Biology of Lomonosov Moscow State University, Moscow, Russia*
[5]*Linköping University, Linköping, Sweden*
[6]*Riga Stradins University, Riga, Latvia*
[7]*M.P. Chumakov Institute of Poliomyelitis and Viral Encephalities, Russian Academy of Sciences, Moscow, Russia*

Correspondence should be addressed to Anastasia Latanova; aalatanova@gmail.com

Received 28 December 2016; Accepted 13 April 2017; Published 22 June 2017

Academic Editor: Masha Fridkis-Hareli

Reverse transcriptase (RT) is a key enzyme in viral replication and susceptibility to ART and a crucial target of immunotherapy against drug-resistant HIV-1. RT induces oxidative stress which undermines the attempts to make it immunogenic. We hypothesized that artificial secretion may reduce the stress and make RT more immunogenic. Inactivated multidrug-resistant RT (RT1.14opt-in) was N-terminally fused to the signal providing secretion of NS1 protein of TBEV (Ld) generating optimized inactivated Ld-carrying enzyme RT1.14oil. Promotion of secretion prohibited proteasomal degradation increasing the half-life and content of RT1.14oil in cells and cell culture medium, drastically reduced the residual polymerase activity, and downmodulated oxidative stress. BALB/c mice were DNA-immunized with RT1.14opt-in or parental RT1.14oil by intradermal injections with electroporation. Fluorospot and ELISA tests revealed that RT1.14opt-in and RT1.14oil induced IFN-γ/IL-2, RT1.14opt-in induced granzyme B, and RT1.14oil induced perforin production. Perforin secretion correlated with coproduction of IFN-γ and IL-2 ($R = 0{,}97$). Both DNA immunogens induced strong anti-RT antibody response. Ld peptide was not immunogenic. Thus, Ld-driven secretion inferred little change to RT performance in DNA immunization. Positive outcome was the abrogation of polymerase activity increasing safety of RT-based DNA vaccines. Identification of the molecular determinants of low cellular immunogenicity of RT requires further studies.

1. Introduction

Starting from the first DNA immunization in 1991, multiple gene-based HIV vaccines have undergone preclinical and clinical trials [1–3]. Several of them that initially aimed to induce strong T cell responses failed to do so indicating a necessity to optimize both genes and their combinations. Several preclinical and clinical studies employed HIV *pol*

gene, full-length or in fragments [4–6]. Plasmids encoding some of the *pol* gene products, as protease and integrase, were shown to be immunogenic in both preclinical and clinical trials [7–10]. At the same time, numerous trials showed an impaired immunogenicity of HIV-1 reverse transcriptase (RT) [11–13]. A recent study by Garrod et al. compared the performance in C57BL/6 mice of DNA vaccines encoding single HIV antigens in combination with HIV gag- and pol-based DNA immunogens. The efficacy of vaccination was tested by challenge with a chimeric EcoHIV virus that can infect mice [14]. At 60 days, there was significantly lower frequency of induced antigen-specific CD8($^+$) T cells in the spleens of pCMVgag-pol-vaccinated mice compared with mice immunized with single pCMVgag. Furthermore, while short-term viral control of EcoHIV was similar for gag- and gag-pol DNA-vaccinated mice, only gag DNA-vaccinated ones were able to control EcoHIV two months postvaccination, indicating that inclusion of the HIV *pol* gene may reduce the durable control over viral replication [14].

HIV enzymes encoded by *pol* gene, including RT, are crucial if aiming at immunotherapeutic vaccination which would prevent drug resistance in HIV infection [15]. Potent immunogenic performance of all three HIV enzymes is a prerequisite of the efficacy of such immunotherapy. We and others performed series of studies aimed to improve the immunogenicity of RT, a key enzyme determining HIV-1 resistance to antiretroviral therapy, but with a limited success [12, 13, 16–18]. Lately, we found that cells expressing HIV-1 RT produce reactive oxygen species (ROS) and express high levels of phase II detoxifying enzymes that interfere with the immune response against this enzyme [19, 20]. Oxidative stress is induced by a wide panel of RT variants, drug-resistant, expressed from viral and expression-optimized genes, enzymatically active and inactive [19] indicating that the ability to induce oxidative stress and oxidative stress response is a property of a domain (domains) within the protein rather than the consequence of its enzymatic activity. We hypothesized that cellular immunogenicity of HIV RT in DNA immunization may be increased by decreasing the levels of this stress-inducing protein in the expressing cells.

We tested if this is the case by artificially promoting RT export. For this, we provided a multidrug-resistant variant of HIV-1 RT (RT1.14) [16], complemented for safety sake, with mutations inhibiting polymerase and RNase H activity, with a leader signal peptide (Ld) of the nonstructural protein 1 of tick-borne encephalitis virus (NS1 of TBEV). NS1 is synthesized as a monomer and dimerizes after the posttranslational modification; it is also expressed on the cell surface and is secreted as a hexamer [21–23]. Ld peptide is responsible for the presentation of NS1 on the cellular surface and further secretion [24–26]. We characterized the properties of Ld-RT1.14 chimera such as the half-life, route of degradation, efficacy of secretion, capacity to induce oxidative stress, and oxidative stress response and, finally, studied its performance in DNA immunization in a mouse model. Retargeting of RT to ER with subsequent secretion resulted in an increase in the RT expression levels due to protein stabilization and also, interestingly, in the nearly complete inhibition of the residual polymerase activity retained in the inactivated RT. Secretion led to a mild reduction of oxidative stress, but no significant enhancement of the cellular immune response in the experimental DNA immunization. Immune response to RT remained tinted towards the production of RT-specific antibodies, typical to M2 polarization of macrophages and Th2 polarization of T-cells in the settings of oxidative stress [27–29].

2. Materials and Methods

2.1. Plasmids. Expression-optimized gene encoding reverse transcriptase derived from the patient infected with multidrug-resistant HIV-1 clade B isolate (RT1.14, [18]) with mutations D185N, D186N, and E478Q abrogating polymerase and RNase H activities was cloned into pVax1 vector (Invitrogen, USA) generating pVaxRT1.14opt-in which was described by us earlier [19]. Sequence encoding RT1.14opt-in was used as a backbone to design a chimeric gene encoding RT1.14 with the N-terminal insertion of the leader sequence of NS1 protein of TBE. The latter was cloned from plasmid pLdNS1 carrying the gene of NS1 protein of Western European subtype of TBE [24]. The cloned fragment encoded 25 amino acids of a leader sequence of NS1 and the N-terminal methionine for an effective initiation of translation. This fragment including 78 nucleotide b.p. was amplified with KAPA HiFi Hot Start DNA polymerase (Kapa Biosystems, USA) and Ld-NheI-F (5′-ATA-CGC-AAG-CTA-GCA-ATA-TGA-GAA-ACC-CTA-CAA-TG-3′) and Ld-BamHI-R (5′-TCT-AAC-AGG-ATC-CCG-CCC-CCA-CTC-CAA-GGG-3′) primers (Syntol, Russia), containing restriction sites *NheI* and *BamHI*, respectively. Resulting PCR product was cloned at the 5′-terminus of RT1.14 gene using *NheI* and *BamHI* restriction sites. Fusion of NS1 Ld- and RT1.14opt-in-encoding fragments formed one open-reading frame encoding optimized inactivated leader-fused RT1.14, dubbed RT1.14oil, within the plasmid pVaxRT1.14oil. Nucleotide sequence of the cloned fragment was confirmed by sequencing. For immunization, plasmids were purified using Plasmid Endofree kits (Qiagen, Germany) as described by the manufacturer.

2.2. Expression of Homo- and Heterodimers of Reverse Transcriptase in E. coli. For expression in *E. coli* genes of active and inactivated RT, HIV-1 HXB2 and RT1.14 were cloned into a two-cistron vector pET-2c. Proteins were expressed in *E. coli* and purified by ion exchange chromatography as homodimers using the protocol described earlier [30]. Wild-type p66/p51 heterodimeric HIV-1 RT was expressed in M-15 [pREP4] *E.coli strain* transformed with the plasmid p6HRT [31] and purified as previously described [32].

2.3. RT Polymerase Activity Assay. The RT assays using activated DNA were performed as follows: the standard reaction mixture (20 μl) contained 0,75 μg of activated DNA, 0,02–0,05 μg RT, 3 μM dATP, 30 μM of dCTP, dGTP, and dTTP, 1 μCi [a-32P]dATP in a Tris-HCl buffer (50 mM, pH 7,5) containing also 10 mM $MgCl_2$, and 0,2 M KCl. The reaction mixtures were incubated for 12 minutes at 37°C and applied onto Whatman 3MM filters with a 0,5 M EDTA solution to

stop reaction. After drying on air, the filters were washed twice with 10% trichloracetic acid, then twice with 5% trichloracetic acid, and once with ethanol and dried on air. The radioactivity was measured by the Cherenkov method [33]. Radioactivity was determined using a Tri-Carb 2810 TR scintillation counter (Perkin Elmer, USA).

2.4. RNase H Activity Assay. RNase H activity assay of HIV-1 RT was tested by using 6,7 mmol of 18-ribo-Fl/18-deoxy duplex (18-mer oligoribonucleotide (18-ribo-Fl: 5′-r(GAUCUGAGCCUGGGAGCU)-fluorescein-3′) and 18-mer oligodeoxyribonucleotide (18-deoxy d5′-d(AGCTCCCAGGCTCAGAUC)-3′)). RNA/DNA duplex was added to reaction mixtures consisted of 15 μL of 50 mM Tris/HCl (pH 8,0), containing 60 mM KCl, 10 mM $MgCl_2$, and various concentrations of p66/p51 RT variant (5, 20, 100, and 400 nM) followed by a 15 min incubation at 37°C. The reaction was stopped by adding 80 μl of 7 mM EDTA, 0,375 M sodium acetate, 10 mM Tris-HCl (pH 8,0), and 0,125 mg/ml glycogen. The mixture was extracted by phenol/chloroform, and RNA/DNA fragments were precipitated with ethanol. The reaction products were separated by electrophoresis in a 20% polyacrylamide/7 M urea gel (PAGE). Gel images were recorded using Typhoon FLA 9500TM Phosphorimager (Molecular Dynamics, Israel) and then quantified using Quantity One 4.6.6. (Bio-Rad, USA). The experiments were conducted three times.

2.5. Eukaryotic Cell Cultivation and Transfection. HeLa and HEK293T cells were cultivated at 37°C in 5% CO_2 in DMEM medium (PanEco, Russia) supplemented with 10% of fetal bovine serum (FBS; HyClone, USA) and 100 μg/ml of streptomycin/penicillin mixture. Transfection was performed with Lipofectamine LTX (Invitrogen, USA) in accordance with manufacturer's instructions. Cells were grown for 48 hours, cells were harvested, and cell culture fluids were collected and used for further analysis of RT expression, studies of oxidative stress, and RT polymerase activity.

2.6. Immunostaining of RT Expressing Eukaryotic Cells. HeLa cells were grown and transfected on the cover glass (20 × 20 mm) in the 6-well cell culture plates (Corning, Costar, USA). Two days posttransfection, cells were fixed with methanol-acetone (1 : 1). After fixation, cells were incubated in the staining buffer (PBS, 2% BSA, 0,2% Tween 20, and 10% glycerol) containing monoclonal murine anti-RT (1 : 10) or polyclonal rabbit anti-RT (1 : 100) and subsequently stained with FITC-conjugated anti-murine or anti-rabbit antibodies (1 : 50) (Dako, Denmark) as described previously [12]. Cell nuclei were stained with DAPI fluorochrome (Invitrogen, USA). For surface protein analysis, cells were fixed with 2% paraformaldehyde, then blocked with 5% BSA in PBS, and incubated with anti-RT antibodies diluted in the staining buffer without Tween 20. Slides were read on the confocal microscope Leica TCS5 (Leica, Germany).

2.7. Quantification of RT Expression in Eukaryotic Cells and Cell Culture Fluids by Western Blotting. RT1.14-expressing HeLa were lysed and analyzed by Western blotting using polyclonal rabbit anti-RT [34] or monoclonal murine anti-RT antibodies [35] as primary and anti-rabbit or anti-mouse HRP-conjugated antibodies as secondary (Jackson, USA) as was previously described [12]. Cell culture fluids were centrifuged at 5000 ×g and then concentrated approximately 10 times with Vivaspin 500 units (Sartorius Stedim, Germany) with a 30,000 MWCO membrane. Concentrated culture fluid was mixed with Laemmli buffer and then subjected to Western blotting same way as the cell lysates. Immune complexes on the membrane were detected with ECL (Amersham, USA) and X-ray film (FujiFilm, Japan). The data was processed in ImageJ software (http://rsb.info.nih.gov/ij). After RT-specific staining, blots were washed and restained first with monoclonal anti-β-actin murine antibodies (Sigma, USA) and then with anti-mouse HRP-conjugated antibodies (Dako, Denmark). To assess the level of RT expression per cell, the percent of cells expressing RT was estimated from the efficacy of transfection established in a control cotransfection with GFP plasmid (peGFP-N1, Novagen, Germany) used as a reporter. The number of cells for each sample was counted with hemocytometer and certain number of cells was taken for Western blotting analysis. The number of transfected cells was estimated from the percentage of transfection efficacy. The amount of RT protein in the lysed cells and culture fluids was calculated based on a standard curve built using the recombinant RT 1.14 protein and dispensed in serial dilutions in the concentration range from 1 to 20 ng per well; the latter samples were analyzed together with the lysates as described earlier [12]. RT content per cell was calculated by dividing these values by the number of transfected cells.

2.8. Quantification of RT Polymerase Activity in Cell Lysates and Cell Culture Fluids. Two days posttransfection of HeLa cells with pVaxRT1.14opt-in or pVaxRT1.14oil plasmids, cell culture fluids were collected and concentrated as described above. Cells were lysed in TNEV buffer containing 50 mM Tris HCl (pH 7,5), 1% Triton X-100, 2 mM EDTA, and 100 mM NaCl supplemented with protease inhibitor cocktail (Sigma, USA). RT activity in the cell lysates and cell culture fluids was assessed by Lenti RT activity kit (Cavidi, Sweden) following instructions of the manufacturer. The amount of protein determined by the polymerase activity assay was normalized to the total amount of RT protein assessed by Western blotting using the calibration curve built with the recombinant RT1.14 protein in a concentration range from 1 to 20 ng per well. The data obtained represented the relative polymerase activity of the samples.

2.9. Measurement of Reactive Oxygen Species. Measurement of reactive oxygen species was performed as described earlier [36]. In brief, 40 hours posttransfection, HEK293T cells were incubated for 30 min in cell culture medium containing 25 μM 2′,7′ dichlorodihydrofluorescein diacetate (DCFH). Fluorescence intensities were measured using Plate CHAMELEON V reader (Hidex Ltd., Finland) with the excitation at 485 nm and emission at 535 nm.

2.10. Reverse Transcription and Quantitative PCR (RT-qPCR). RNA was isolated from 5×10^5 transfected HEK293T cells with

PerfectPure RNA Cultured Cell kit (5 Prime, Germany, USA) and then transcribed with Reverse Transcription System (Promega, USA) with random hexamer primer according to manufacturer's protocol. RT-qPCR was performed using iQ5 Real-Time PCR Detection System (BioRad, USA) and primers and probes which were described earlier [36]. A standard reaction mixture (50 μl) contained Taqman primer/probe combination, cDNA equivalent to 100 ng of total RNA, and qPCR-HS master mix. The thermal conditions for PCR reaction for all genes were 55°C for 5 min, 95°C for 10 min followed by 40 cycles at 95°C for 10 sec, and 57°C for 1 min (signal collection temperature). Relative quantitative analysis was performed by comparing threshold cycle number for target genes and a reference β-actin mRNA, amplified in separate tubes.

2.11. Degradation of RT in Eukaryotic Cells. The measurement of half-life of the protein was done using the cycloheximide chase assay based on the method described earlier [37]. For this, HeLa cells 48 hours posttransfection were treated with cycloheximide (Sigma-Aldrich, USA) at the final concentration 100 μg/ml. After 0, 2, 4, and 6 hours of incubation, the cells were harvested, lysed, and analyzed by Western blotting. The half-life time for the protein was calculated with a standard formula: $T_{1/2} = -0{,}693\ t/\ln N/N_0$, where N_0 is the initial amount of protein and N the amount of protein at time t. To evaluate the role in RT degradation of the proteasome and lysosome inhibitors, HeLa cells were transfected with pVaxRT1.14opt-in or pVaxRT1.14oil and 24 hours posttransfection treated with the cellular protease inhibitors at the final concentrations indicated in the brackets: E64 (10 μM), leupeptin (10 μg/ml), aprotinin (10 μg/ml), pepstatin A (7,5 μM), MG132 (5 μM), or epoxomicin (0,1 μM) (all from Calbiochem, USA). After 18 hours of incubation with the inhibitors, cells were lysed and the residual amount of RT was estimated by Western blotting.

2.12. DNA Immunization of Mice. Eight-week-old BALB/c mice (8 weeks, Charles River Laboratories, Sandhofer, Germany) were housed under a light-dark (12 h/12 h) cycle with ad libitum access to water and food. Animals were anesthetized by a mixture of 4% isoflurane with oxygen and maintained in 2,3% isoflurane flow administered through facial masks during all intradermal injections and electroporation. All experimental procedures were approved by the local ethical committee. Groups consisting of 4–6 mice were injected with 10 μg of pVaxRT1.14opt-in or pVaxRT1.14oil, or with empty pVax1 vector, mixed with equal amount of pVaxLuc encoding firefly luciferase administered with 29G insulin needles in 20 μl of PBS. Each mouse got two intradermal DNA injections, to the left and to the right of the base of the tail. Immediately after, injection sites were subjected to electroporation using DermaVax DNA vaccine delivery system (Cellectis Glen Burnie, France) as described earlier [38]. In day 21 postimmunization, mice were bled and sacrificed and their spleens were collected.

2.13. Analysis of Cellular Immune Response

2.13.1. INF-γ and IL-2 Fluorospot. The spleens of immunized mice were homogenized, and the splenocytes were isolated as described in [8]. The cells were incubated in RPMI medium supplemented with 2 mM L-glutamine, 100 μg/ml of streptomycin/penicillin mixture (all from Sigma-Aldrich, USA), and 10% FBS (Gibco, Invitrogen, USA) (complete media), with the following antigens taken in concentration 10 μg/ml: synthetic peptides corresponding to RT aa375-389 (ITTESIVIWGKTPKF), 465-476 (KVVPLTNTTNQK), 514-528 (ESELVNQIIEQLIKK), 528-543 (KEKVLAWVPAHKGIG), leader sequence of NS1 (RNPTMSMSFLLAGGLVLAM TLGVGA) (all peptides from GL Biochem Ltd., China) and RT 1.14 protein. Concanavalin A (ConA, 5 μg/ml) was used as a positive control and cell culture medium as a negative control. After 20 hours of incubation, IFN-γ and IL-2 secretion by splenocytes was assessed by dual IFN-γ/IL-2 Fluorospot (Mabtech, Sweden) in accordance with the protocol provided by the manufacturers. The number of cells secreting cytokines was calculated with fluorimeter AID ELISpot (Autoimmun Diagnostika GmbH, Germany).

2.13.2. Detection of Secreted Granzyme B and Perforin by Splenocytes. Isolated murine splenocytes were applied on 96-well V-plates in 200 μl RPMI-10% FBS and stimulated in duplicates with the recombinant RT1.14 protein and RT peptides, same as used in Fluorospot test (10 μg/ml). Cells were incubated for 3 days at 37°C and 5% CO_2. After that, 100 μl of cell culture fluid from each well was collected, duplicates were pooled, redivided into two 100 μl aliquots, and subjected to analysis by sandwich ELISA for granzyme B (DuoSet Development kit; R&D Systems Europe Ltd.) and perforin (PRF1; Hölzel Diagnostika Handels GmbH, Germany). Kits were used as recommended by the manufacturers.

2.14. Detection of Anti-RT Specific Antibodies by Indirect ELISA. 96-well microtiter plates (Nunc Maxisorp, Denmark) were coated with recombinant RT1.14 protein or NS1 Ld peptide diluted in PBS at 0,3 μg/ml by overnight incubation at 4°C. Solution was discarded, and plates were washed with PBS containing 0,05% Tween 20. Individual mice sera diluted in HIV-scan buffer (2% normal goat serum, 0,5% BSA, 0,05% Tween 20, and 0,01% sodium merthiolate) in five (IgA sybtype) or three (IgG subtypes) fold-steps starting from 10 (IgA) or 200 (IgGs) were applied on the plates and incubated overnight at 4°C. The plates were washed as above, and HRP-conjugated goat anti-mouse IgG or IgA (Sigma, USA) diluted in HIV-scan buffer was applied and incubated for 1,5 hours at 37°C. After the incubation, plates were washed as above and color reaction was developed with 3,3′,5,5′- tetramethylbenzidine substrate solution (TMB; Medico-Diagnostic Laboratory, Russia). Reaction was stopped by addition of 50 μl of 2,5 M of sulfuric acid, and optical density was recorded at a dual wavelength of 450 and 650 nm. The cutoff for specific RT antibody response was set as mean OD values showed by sera of vector-immunized mice at this time point +3 SD. For positive sera showing OD values exceeding the cutoff, endpoint dilution titers were established from the titration curves.

2.15. In Vivo Monitoring of Reporter Expression. Bioluminescence emission from the area of DNA-immunogen/Luc gene

injection was analyzed days 1, 3, 9, 15, and 22 postimmunization. To estimate luciferase expression in vivo, mice were intraperitoneally injected with the solution of D-luciferin (Perkin Elmer, USA) in PBS in 15 mg/ml at a dose of 100 μl for 10 g of body weight and left for 5 minutes. In vivo imaging of bioluminescence was performed with a highly sensitive CCD camera, mounted in a light-tight chamber (Spectrum CT, Perkin Elmer, USA). Anesthesia was induced by 4% isoflurane and maintained by 2,3% isoflurane throughout the imaging procedure. Camera exposure time was automatically determined by the system and varied between 1 and 60 s depending on the intensity of the bioluminescent signal. Regions of interest were localized around the injections sites and were quantified as the total luminescence flux in photons/s. CT/BLI data were processed using the Living Image® software version 4.1 (Perkin Elmer, USA) to generate values of total photon flux and mean photon flux (photons/sq cm) from the injected area.

2.16. Statistics. Statistical evaluations were done using STATISTICA AXA 10.0 (StatSoft Inc., OK, USA). Nonparametric statistics was chosen as appropriate for sample sizes <100 entries. Continuous but not normally distributed variables, such as the average radiance in photons/s/cm^2/Sr, antibody titers, or the number of cytokine-producing SFCs, cytokine levels in pg/ml, were compared in groups by the nonparametric Kruskal-Wallis and pairwise by Mann–Whitney U test. Correlations were run using the Spearman rank order test. p values <0.05 were considered significant.

3. Results

3.1. Design and Expression of the Inactivated Multidrug Resistant RT Fused to the Leader Peptide of TBEV NS1 (RT1.14oil). The secretable RT chimera was designed based on the plasmid pVaxRT1.14opt-in [19] which encodes an inactivated multidrug-resistant RT variant (RT1.14opt-in). Original RT1.14 amino acid sequence originated from HIV-1 clade B isolate from patient with multiple drug resistance [18] modified by mutations introducing two D/N substitutions in the polymerase YMDD motif, and one E/Q substitution in the RNase H DEDD motif, shown earlier to inhibit respective enzymatic activities [39–42]. Inactivation of the polymerase and RNase H moieties of RT was confirmed in in vitro assays done on the "classical" RT derived from HIV-1 HXB2 strain and its analogue inactivated by D185N, D186N, and E478Q mutations, represented, for adequate characteristics of the enzymatic activity, by p51/p66 heterodimers (Supplementary Figure S1 available online at https://doi.org/10.1155/2017/7407136). Introduction of D185N and D186N mutations completely abrogated the polymerase activity of the enzyme (Figure S1A). We have also assessed the RNase H activity of the active and inactivated RTs (Figure S1B, C). The wild-type RT demonstrated significant activity in a broad range of enzyme concentrations tested (Figure S1B, C). The efficacy of hydrolysis of the 18-mer RNA/DNA hybrid duplex strongly depended on the enzyme concentration (Figure S1B, C). On the contrary, RT inactivated with mutation E478Q in the RNase H moiety was practically inactive even at high enzyme concentrations; the efficiency of hydrolysis was reduced by over 90% compared to the wild-type enzyme. These results confirmed the adequacy of inactivation strategy used to generate the inactivated drug-resistant HIV RT.

To design the secreted form of inactivated RT1.14, we used a specialized signal of the heterologous viral protein, namely of the nonstructural protein 1 of tick-borne encephalitis virus (TBEV NS1) [24] (Figure 1). The leader signal of TBEV NS1 (Ld) consisting of the last 25 amino acid residues of TBEV envelope protein (glycoprotein) E is responsible for the transport to ER and secretion of the downstream NS1 protein [25, 26]. DNA sequence encoding LdNS1-RT1.14opt-in chimera was generated by PCR and cloned into pVax1 vector yielding a plasmid dubbed pVaxRT1.14oil (Figure 1).

We have then compared the expression profiles of the parental inactivated RT1.14 and of Ld-RT1.14 chimera. For this, we transfected HeLa cells with pVaxRT1.14opt-in and pVaxRT1.14oil plasmids; 48 hours posttransfection, cell lysates and cell culture fluids were collected and analyzed by Western blotting. Lysates of cells from both transfections contained proteins with molecular mass of 66 kDa coinciding with the expected molecular mass of RT1.14 [43] (Figure 2(a)). The amount of the 66 kDa product detected in the lysates of cells transfected with pVaxRT1.14oil was approximately 1,4 times higher than in the lysates of cells transfected with the parental RT1.14opt-in gene (Figure 2(a)). Both lysates contained also a protein with the molecular mass of 51 kDa corresponding the polymerase subunit of HIV-1 RT [44]. Besides, lysates of cells transfected with pVaxRT 1.14oil contained products of higher molecular mass (over 170 kDa) that were specifically stained with anti-RT antibodies; suggestively, the aggregates formed as a result of RT overexpression. RT1.14opt-in was found in the cellular fraction and, interestingly, in equal amounts, in the cell culture fluid from the expressing cells (Table 1). Ld-RT1.14 chimera encoded by pVaxRT1.14oil was preferentially (>70%) found in the cell culture fluid (Table 1).

3.2. Enzymatic Activity of RT in the Expressing Cells and in the Culture Fluids. Next, we analyzed the lysates of cells transfected with RT1.14opt-in and RT1.14oil genes for the polymerase activity of RT by an ELISA-based test (Cavidi, Sweden). We also collected, concentrated, and subjected to the HIV RT activity test samples of the cell culture fluids. RT protein content in these probes was estimated by Western blotting.

A residual RT polymerase activity was detected in all probes, but not in the probes from nontransfected HeLa cells. We evaluated the specific RT activity per 10^5 of HeLa cells transfected with RT1.14opt-in and with RT1.14oil coding plasmids. Lysates of 10^5 cells expressing RT1.14opt-in had RT activity corresponding to 9,8 pg of the active protein, that is, retained 0,00012% activity (9,8 pg of the active form in the total 8,3 μg of the enzyme by Western blotting; Table 1). Lysates of 10^5 cells expressing RT1.14oil retained 0,00005% of the activity (5,7 pg of the active form in the total 11,3 μg of the enzyme; Table 1). At the same time, cell culture fluids of RT1.14opt-in expressing cells retained 0,00053% of the activity (74,4 pg out of 14 μg), which is four times more than

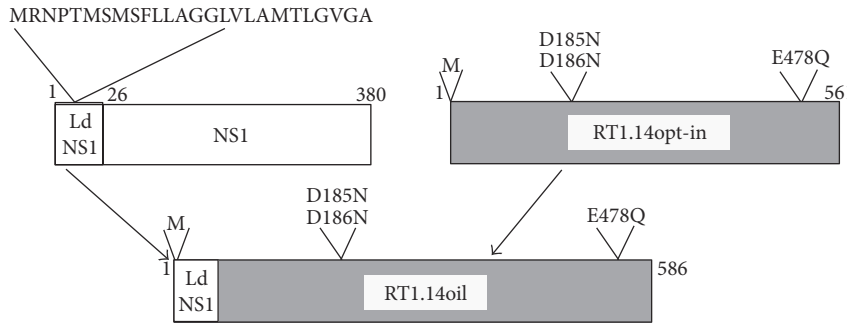


Figure 1: Schematic representation of chimeric drug-resistant HIV-1 reverse transcriptase carrying the N-terminal insertion of the N-terminal signal peptide of NS1 protein of TBEV (RT1.14oil). Rectangular boxes stand for polypeptide chains of NS1 with leader signal peptide (white) and RT1.14opt-in sequence (gray) with amino acid substitutions in polymerase (D185N, D186N) and RNase H signature motives (E478Q) leading to inactivation of respective enzymatic activities in the resultant RT1.14oil polyprotein. Amino acid positions in the polypeptide chain are designated with numbers.

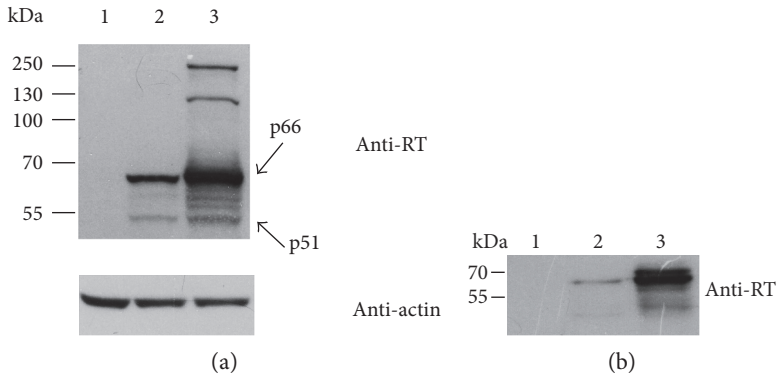


Figure 2: Expression of HIV-1 RT1.14 fused to the N-terminal signal peptide of NS1 protein of TBEV in HeLa cells resolved by SDS-PAGE. (a) Western blotting of the lysates of HeLa cells transfected with pVax1 (1), pVaxRT1.14opt-in (2), and pVaxRT1.14oil (3) plasmids; upper panel, membranes stained with anti-RT; lower panel, with anti-β-actin antibodies for signal normalization. (b) Western blotting of the culture fluids of cells transfected with pVax1 (1), pVaxRT1.14opt-in (2), and pVaxRT1.14oil (3). Blots were stained with the specific polyclonal anti-RT antibodies as described in the experimental section "Materials and Methods." To control equal loading of the samples of cell lysates on the gel, membranes were for the second time stained with anti-β-actin antibodies. Positions of the molecular mass markers are given to the left.

the specific activity in the cells (Table 1). The latter pointed that RT1.14opt-in is secreted in the active form even without the signal peptide. On contrary, cell culture fluids of RT1.14oil-expressing cells contained only 0,000014% of the active enzyme (4,9 pg per 34,6 μg), that is, 3,5 times less than the lysates of RT1.14oil-expressing cells (Table 1). Thus, RT1.14oil was effectively secreted, but almost exclusively in the inactive form.

3.3. RT Localization in the Expressing Eukaryotic Cells. Intracellular localization of RT variants was investigated by the immunofluorescence microscopy. HeLa cells were transfected with pVaxRT1.14opt-in and pVaxRT1.14oil plasmids and subsequently stained with anti-RT and then FITC-conjugated secondary antibodies. RT1.14opt-in was distributed in cell cytoplasm homogeneously (Figure 3(a)), whereas the distribution of RT1.14oil was more granular (Figure 3(b)). Besides, cultures of the cells transfected with pVaxRT1.14oil plasmid were characterized by the presence of granules specifically stained with anti-RT antibodies that were localized inside and outside the cells and on the cell surface (Figure 3(b), indicated by arrows). To confirm the cell surface localization of the protein, transfected cells were fixed with 2% paraformaldehyde with subsequent staining in the absence of detergent. Surface of the cells transfected with the initial plasmid pVaxRT1.14opt-in was not specifically stained (Figure 3(c)), whereas the surface of pVaxRT1.14oil-transfected cells was specifically stained with anti-RT antibodies (Figure 3(d)). To localize RT in relation to ER, HeLa cells were treated with TRITC-conjugated antibodies to an ER marker calreticulin, and FITC and TRITC signals were overlaid (Figures 3(e) and 3(f)). In the case of RT1.14oil, RT-specific signal was partially colocalized with ER marker and partially with the secretory vesicles, whereas RT1.14opt-in had a more prominent colocalization with ER.

3.4. RT Protein Degradation. Next, we estimated the speed of degradation of RT1.14oil in HeLa cells and investigated the contribution of cellular proteases into its proteolysis. To estimate the half-life of the protein, cells transiently expressing

Table 1: Expression, secretion, and residual enzymatic activity of inactivated forms of HIV reverse transcriptase RT1.14opt-in and RT1.14oil in transiently transfected HeLa cells. RT content was determined by Western blotting using protein calibration curves with signal quantification by ImageJ (dubbed WB, in $\mu g/10^5$ cells) and by quantification of enzymatic activity against standard RT preparation using Lenti RT activity test (Cavidi, Sweden) (dubbed EA, in $pg/10^5$ cells). RT content in cell lysates is dubbed "in cells" and in cell culture fluids "secreted." Specific activity was measured as a ratio of protein content evaluated using the enzymatic activity assay to RT content determined by Western blotting.

RT variant	RT content by WB, $\mu g/10^5$ cells			RT content by EA, $pg/10^5$ cells			Specific RT activity (EA/WB), μU	
	In cells	Secreted	Total	In cells	Secreted	Total	In cells	Secreted
RT1.14opt-in	8,30	14,00	22,30	9,80	74,40	85,00	1,18	5,30
RT1.14oil	11,30	34,60	46,00	5,70	4,90	11,00	0,51	0,14

RT1.14oil, or RT1.14opt-in, were after two days treated with cycloheximide and lysed 0, 2, 4, and 6 hours after the on-start of the treatment. The amount of RT1.14opt-in and RT1.14oil in the probes was detected by Western blotting (Figure 4(a)). The half-life of RT1.14opt-in was estimated as 2 and of RT1.14oil as more than 8 hours (approximately 15 hours) which characterizes it as a long-living protein (Figure 4(b)).

For a detailed study of RT proteolysis, on the next day after the transfection, cells were treated with the inhibitors of proteasomal and lysosomal proteolysis. The role of proteasome in degradation was assessed using reversible (MG132) and irreversible (epoxomicin) proteasomal inhibitors. Lysosomal proteolysis was blocked with the inhibitors of cysteine (E64, leupeptin), serine (aprotinin, leupeptin), and acid proteases (pepstatin A). Cells were incubated with the inhibitors for 18 hours and lysed, and RT content was estimated by Western blotting in comparison with the content in the untreated samples (Figure 5(a)). Treatment with MG132 increased RT1.14opt-in content 2,5, and with epoxomicin, 3 times compared to untreated samples; lysosomal inhibitors had no effect. None of the inhibitors had any effect on the intracellular level of RT1.14oil (Figures 5(b) and 5(c)).

3.5. Induction of Oxidative Stress and Oxidative Stress Response in RT1.14-Expressing Cells. Next, we have evaluated the level of oxidative stress induced by RT1.14opt-in and RT1.14oil in HEK293T cells transiently transfected with their genes. Oxidative stress was measured by the formation of reactive oxygen species (ROS) visualized in the presence of fluorogenic 2, 7′-dichlorofluorescein. The presence of both RT variants in cells led to the formation of ROS (Figure 6(a)). Interestingly, secretion resulted in a weak but significant decrease in the levels of ROS production (15%; $p = 0{,}013$; F test; Figure 6(a)). Next, we have studied the induction of oxidative stress response, namely, the induction of expression of phase II detoxifying enzymes, heme oxygenase 1 (HO1), and NAD(P)H-oxygenase 1 (Nqo1). Level of transcription of their genes was evaluated by RT-PCR. Expression of RT variants led to an increase in the levels of expression of both detoxifying enzymes (Figure 6(b)). Cells expressing RT1.14oil demonstrated an increase in the expression of Nqo1 ($p < 0{,}05$; F test; Figure 6(b)) and a tendency to the increased induction of HO1 ($p = 0{,}07$; F test; Figure 6(b)).

3.6. RT1.14 Performance in DNA-Immunized Mice. BALB/c mice were injected with pVaxRT1.14opt-in, pVaxRT1.14oil, or pVax1 plasmids intradermally with subsequent electroporation. On day 22, mice were sacrificed; blood and spleens were collected. Splenocytes were isolated and tested for the ability to proliferate after in vitro stimulation with RT-derived antigens. Immune response was evaluated as the production of IFN-γ and IL-2 by Fluorospot and perforin and granzyme B by sandwich ELISA. RT1.14opt-in- and RT1.14oil-immunized mice demonstrated similar levels of cellular response against the recombinant RT1.14 protein (Figure 7(a)). Immunization with DNA encoding RT1.14opt-in and RT1.14oil induced a weak IFN-γ and IL-2 response against a $CD4^+$ T cell epitope of RT localized at aa 528-543 [45], which tended to be stronger in mice DNA immunized with RT1.14oil (Figure 7(b), $p < 0{,}1$). Peptides representing other known epitopes of RT induced no specific IFN-γ or IL-2 production (data not shown). Interestingly, splenocytes of mice DNA immunized with RT1.14oil responded to stimulation with these peptides by a strong production of perforin (Figure 7(c)). Levels of perforin secretion significantly correlated to the levels of RT-specific production IFN-γ and IL-2 (Figure 7(e); R values >0,8; $p < 0{,}0005$). On contrary, mice DNA immunized with RT1.14opt-in responded to stimulation with RT-derived peptides by a high-level production of granzyme B which was not correlated to either IFN-γ or IL-2 levels (Figure 7(d), data not shown). IFN-γ/IL-2 responses have also been tested in splenocytes stimulated with a peptide encoding a leader sequence of NS1 (NS1 leader peptide). No specific cytokine secretion was detected (data not shown).

Murine sera collected at the endpoint of the experiment were analyzed for anti-RT antibodies by indirect ELISA. No significant difference was detected in antibody titers reached in both groups of RT-immunized mice. Titers of IgG and IgG1 exceeded 50,000 in both groups, IgG2a reached 15,000 in RT1.14opt-in and 6000 in RT1.14oil-immunized mice and IgA 5000 in RT1.14opt-in and 6000 in RT1.14oil-immunized mice, respectively (Figure 8). Ratio of IgG2a/IgG1 was equal to 0,15 in both groups of mice indicating a Th2 type of immune response [46]. TBEV NS1 Ld peptide induced no specific antibodies (data not shown).

3.7. Effector Immune Response to RT and Its In Vivo Evaluation. We have earlier shown that diminishment of

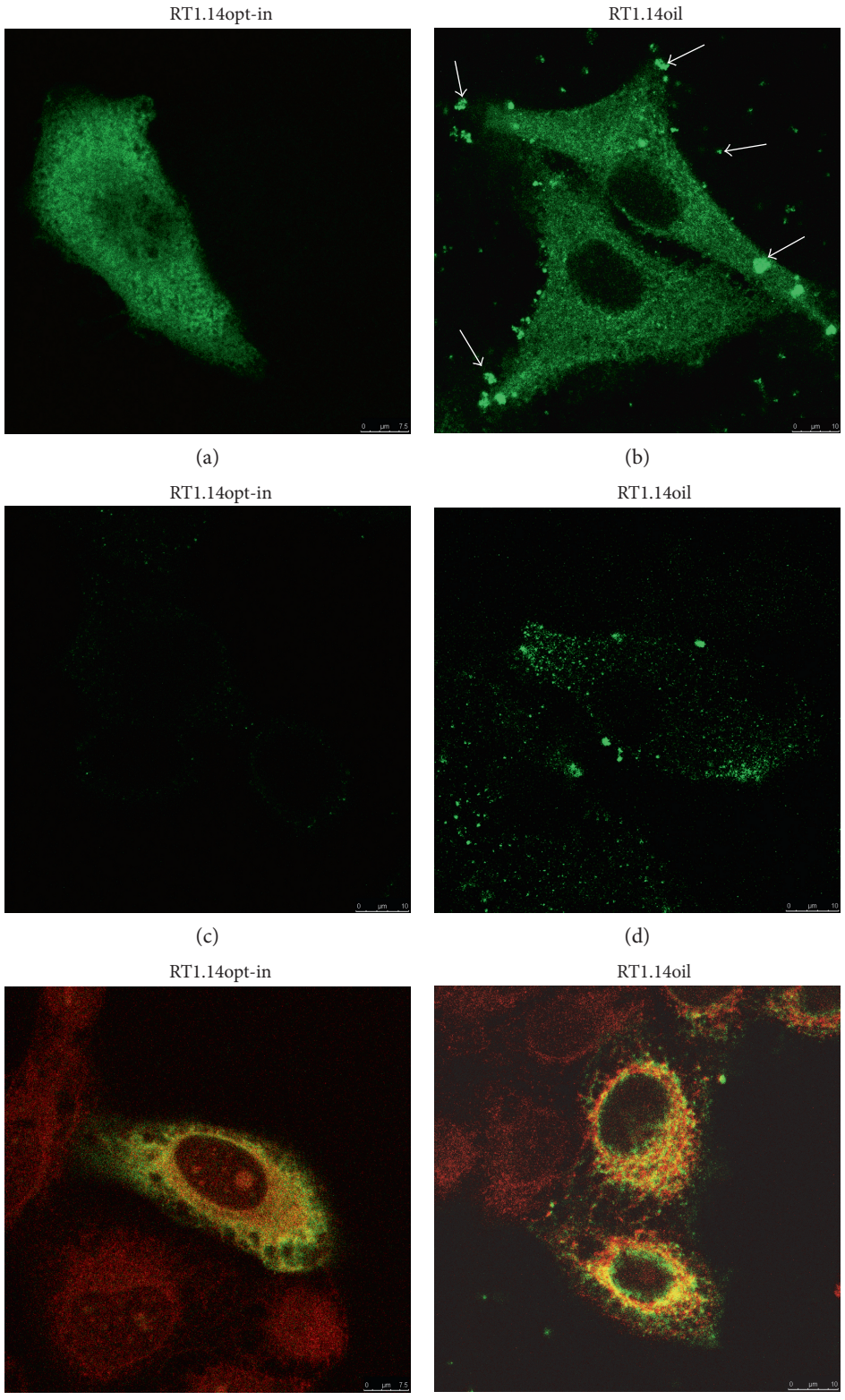


Figure 3: Cellular localization of RT without signal of secretion (RT1.14opt-in) and with a signal peptide (RT1.14oil) in HeLa cells. HeLa cells were transfected with the plasmids pVaxRT1.14opt-in (a, c, e) and pVaxRT1.14oil (b, d, f), grown on slides for 48 hours, fixed with methanol : acetone (1 : 1 *v*/*v*) (a, b, e, f) or 2% paraformaldehyde (c, d), and stained first with polyclonal anti-RT and with secondary FITC-conjugated anti-rabbit antibodies. For localization of RT in relation to ER, HeLa cells were treated with TRITC-conjugated antibodies to an ER marker calreticulin (e, f), and the overlay of FITC and TRITC signals is shown.

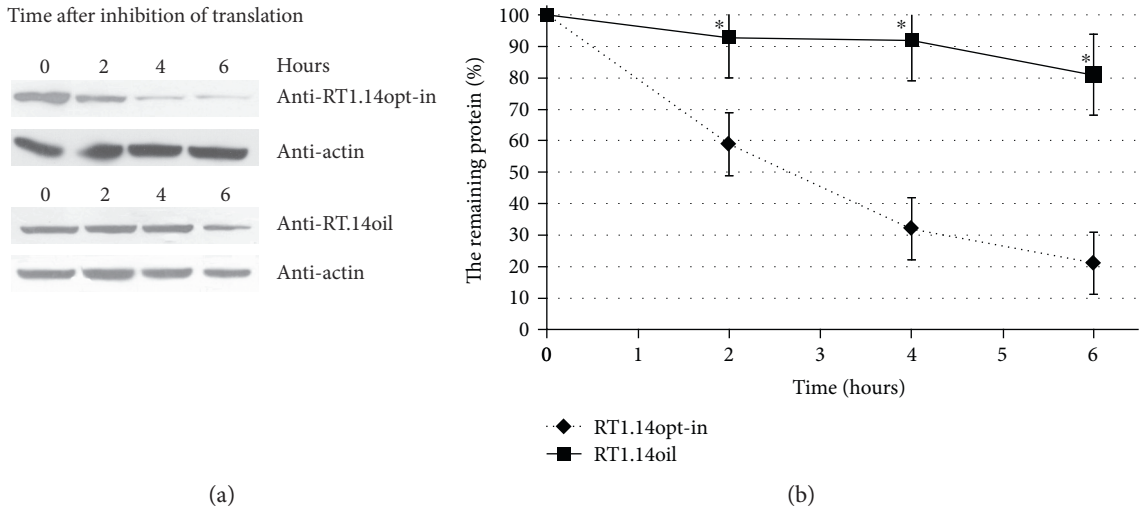


Figure 4: The kinetics of degradation of RT1.14opt-in and RT1.14oil in HeLa cells after blocking of translation with cycloheximide. (a) Western blotting of HeLa cells transfected with pVaxRT1.14opt-in and pVaxRT1.14oil plasmids 0, 2, 4, and 6 hours after cycloheximide treatment. The blots were stained with polyclonal anti-RT antibodies. To control equal loading of the samples on the gel, membranes were restained with anti-β-actin antibodies. (b) Diagram of protein degradation speed. 100% is according to initial amount of the protein. The graph is plotted in accordance with the results of 3 independent experiments, ±SD. $^{*}p < 0{,}05$ (Mann–Whitney test).

bioluminescence induced by the expression of reporter gene coinjected with DNA immunogen correlates with the induction of polyfunctional cellular responses against the immunogen [8, 47] specifically with IFN-γ/Il-2 responses of CD4 T cells [8]. To monitor this effect, mice were coinjected with RT gene variants and a reporter gene encoding firefly luciferase (Luc) and assessed for the levels of bioluminescence on days 1, 3, 6, 9, 15, and 22 postimmunization. We have observed a significant diminishment of in vivo luminescence after day 9 in all immunized mice, independently of the enzyme form (Figure 9).

4. Discussion

Multiple efforts made by us, and by others, to enhance the immunogenicity of HIV-1 RT had limited success [12–14, 16–18]. Recently, we discovered that both the wild-type and drug-resistant RT variants induce oxidative stress in the expressing cells [19]. Oxidative stress has a dual impact on the development of immune response, and low to moderate levels of stress may be stimulating, whereas excessive stress affects both the innate and adaptive immune response [48–51]. High levels of oxidative stress detected in the RT-expressing cells during DNA immunization may suppress the development of specific immune response. We hypothesized that this effect can be neutralized or minimized by an early export of RT from the expressing cells. To test this, we supplemented RT with the leader signal peptide of NS1 protein of TBEV, responsible for cotranslational translocation of the carrier proteins into ER with subsequent secretion [25, 26]. As the backbone for this chimera, we chose the most immunogenic RT variant of the ones we have tested so far, RT of HIV-1 clade B strain with multiple mutations of drug resistance (RT1.14) [18] encoded by the expression optimized gene. RT1.14 was devoid of the enzymatic activities by two D/N mutations in YMDD, and one E/Q mutation in DEDD motives was shown to inhibit the polymerase and RNase H activities, respectively [39–42]. We have proven that these mutations successfully abrogate both activities of the model RT of HIV-1 HXB2 strain. A chimera of the inactivated multidrug-resistant RT with the leader peptide Ld of NS1 dubbed "Optimized Inactivated with Leader sequence" RT1.14, or RT1.14oil, was thus created.

Fusion of the leader peptide sequence changed localization of the chimera. The distribution of RT1.14oil was more granular in comparison to RT1.14opt-in. Also, in the case of RT1.14oil, RT-specific signal was partially colocalized with the ER marker and in part with the secretory vesicles, unlike the parental RT1.14opt-in which had a more prominent ER colocalization. RT1.14oil was detected in the cells and as granules on the cell surface and in the intercellular space. As expected, RT1.14oil was detected also in the cell culture fluids with three times more protein there than in the cells. Altogether, this demonstrated the success of our strategy of RT export. Leader peptide of TBEV NS1 can function as an efficient signal for targeting the heterologous proteins into the secretory pathway. Detection of RT1.14oil both in the cells (cell lysates) and in the cell culture fluids was an expected finding. We have also made an unexpected finding of RT1.14opt-in not only inside the cells (in cell lysates) but also in the cell culture fluids. One of the explanations could be the leakage from apoptotic cells dying due to oxidative stress. However, we have not observed an abnormal/excessive cell death of the RT-expressing cells neither in this nor in the earlier studies [19]. The actual RT export/secretion mechanism requires further elucidation.

Relocalization of RT1.14oil drastically changed its processing. Firstly, its half-life increased from 2 to >8 hours. While RT1.14opt-in was mainly degraded by proteasome, RT1.14oil was insensitive to most of the cellular proteases tested. Thus, retargeting of RT1.14oil into the secretory

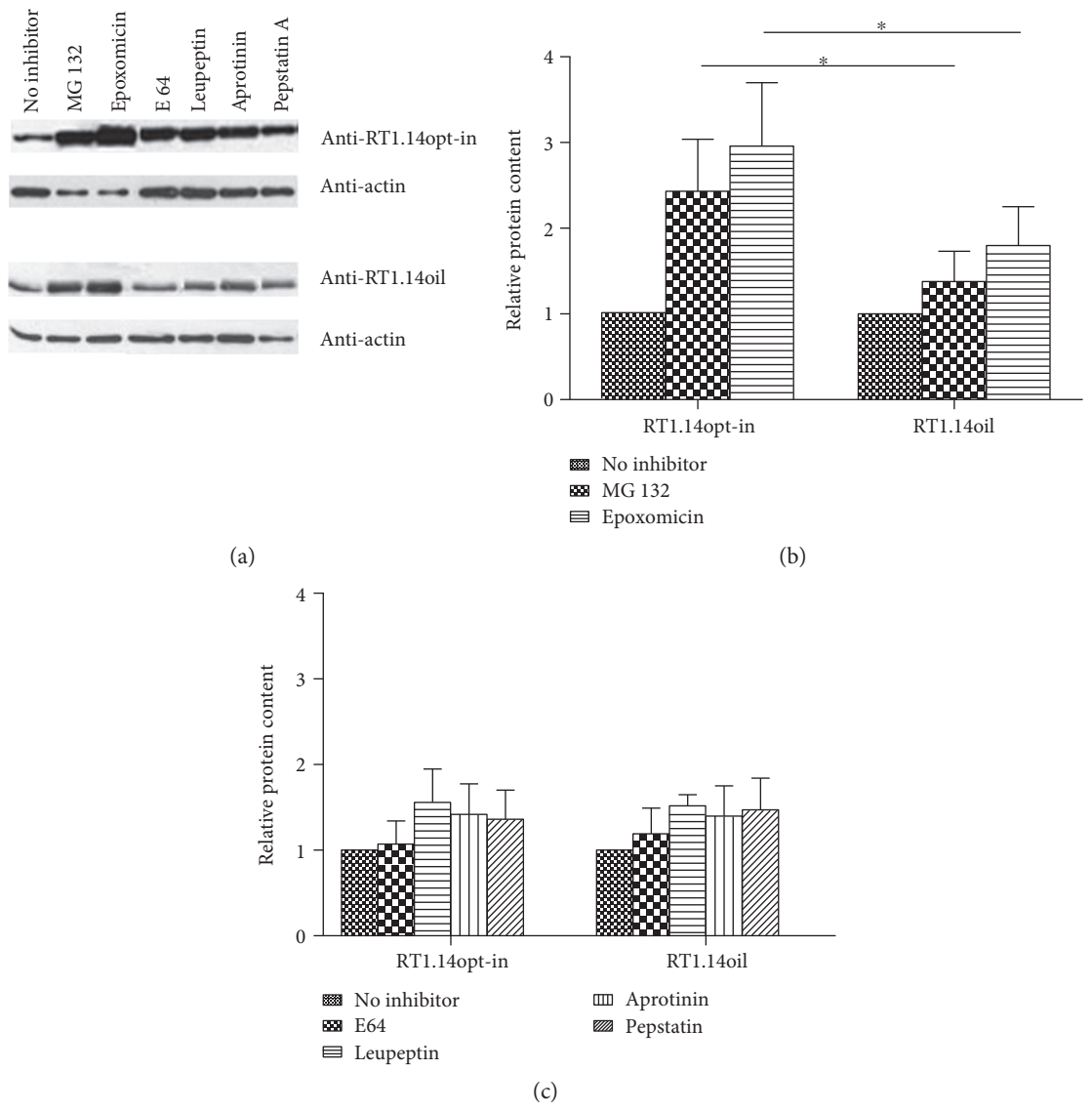


Figure 5: RT accumulation in HeLa cells treated with inhibitors of proteasomal and lysosomal proteolysis. Western blotting of HeLa cells transfected with pVaxRT1.14opt-in and pVaxRT1.14oil after 18-hour incubation with MG132 (5 μM), epoxomicin (0,1 μM), E64 (10 μM), leupeptin (10 μg/ml), aprotinin (10 μg/ml), and pepstatin A (7,5 μM) or without inhibitors. The blots were stained with monoclonal anti-RT antibodies. To control equal loading of the samples on the gel, the membranes were stripped and restained with anti-β-actin antibodies (a). Relative content of RT1.14opt-in and RT1.14oil in the samples treated with proteasomal inhibitors (b). Relative content of RT1.14opt-in and RT1.14oil in the samples treated with lysosomal inhibitors; no difference in response to lysosomal inhibitors between RT1.14opt-in- and RT1.14oil-expressing cells was detected (all p values > 0,05) (c). In (b) and (c), protein content in the untreated samples is taken for 1. Graphs in (b) and (c) represent the results of three independent experiments, +SD. $^*p < 0{,}05$ (Mann–Whitney test).

pathway led to its "escape" from both proteasomal and lysosomal processing which resulted in accumulation of the protein inside and outside of the cells.

By measuring the polymerase activity in the lysates and cell culture fluids of the expressing cells, we found that introduction of the mutations in YMDD motif expected [42, 52] and confirmed here to completely abrogate the polymerase activity of the wild-type HIV-1 RT could not completely inactivate the multidrug-resistant enzyme. Both RT1.14opt-in and RT1.14oil preparations had a residual activity on the level of 0,0005% to 0,00005% of the activity of the equivalent amount of RT of HXB2 strain. Secretion of RT increased the biosafety, as we have observed that RT1.14oil was significantly less active than the parental RT1.14 both in the cells and when secreted in the cell culture medium. One possible explanation is in the RT-induced oxidative stress. Earlier, a phenomenon was described of the superoxide radical-mediated inactivation of an enzyme (myeloperoxidase) secreted by neutrophils [53]. Secreted RT1.14oil can undergo a likewise inactivation. Inactivation can also be due to the protein agglomeration following the accumulation of high amounts of undegradable RT1.14oil protein. We observed granular staining in the cytoplasm of RT1.14oil-expressing cells, and proteins of high molecular mass were specifically stained with anti-RT antibodies in Western blotting of cell lysates, the latter indicating formation of

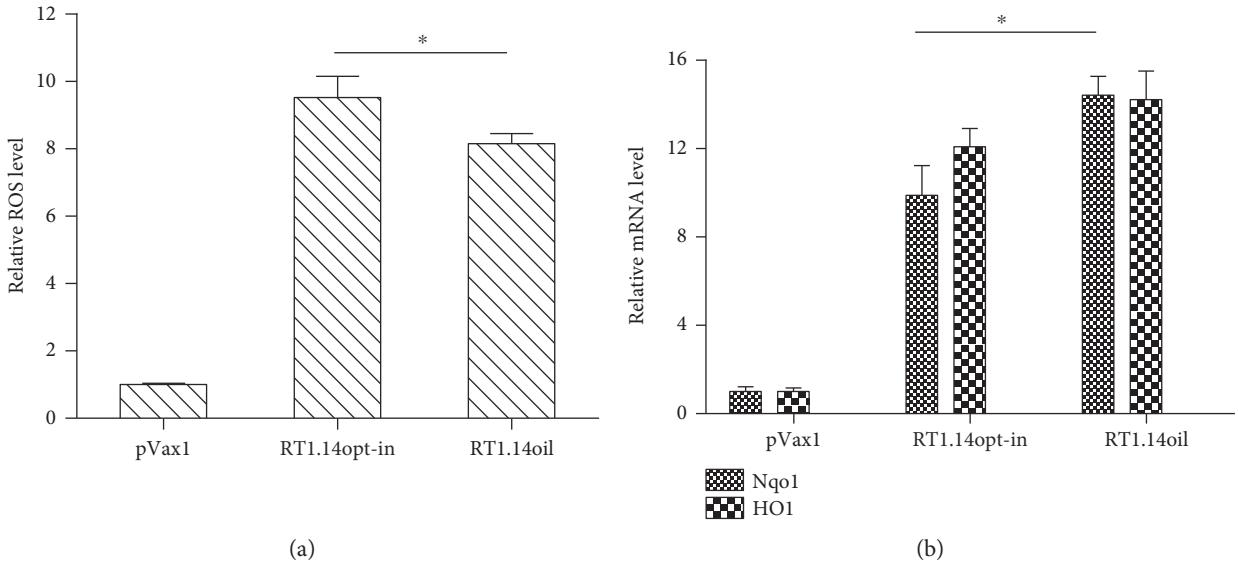


Figure 6: Induction of oxidative stress and oxidative stress response in cells expressing multidrug-resistant inactivated reverse transcriptase RT1.14opt-in and its derivative RT1.14oil carrying a signal peptide. Induction of the oxidative stress was detected as the production of ROS (a) and increase in the levels of mRNA of the phase II detoxification enzymes Nqo1 and HO1 (b). Levels of ROS were normalized to those in HEK293T cells transfected with the empty vector pVax1. Levels of mRNA for Nqo1 and HO-1 were normalized to levels of mRNA for actin and then represented as fold difference to the effect of empty vector pVax1. Data represent the results of two independent experiments, each done in triplicate, +SD. Results are compared using F test (Statistica AXA 10), $^{*}p < 0{,}05$.

cross-linked protein agglomerates, similar to the observations made for the near-infrared reporter proteins [54]. The latter often results in the loss of enzymatic activity (protein function(s)) [55, 56]. Signal peptide-targeted secretion helped to diminish the residual polymerase activity and made RT1.14oil more attractive as the potential DNA vaccine component.

Artificial secretion resulted in a decrease of the capacity of RT to induce oxidative stress with 15% decrease in the production of ROS compared to the parental enzyme. This can be explained by the induction of the ARE-dependent mechanisms of oxidative stress response involving the enhanced expression of phase II detoxifying enzymes, specifically NAD(P)H-oxygenase 1. We have further inquired whether a decrease in the level of oxidative stress induced by RT1.14oil would improve its immunogenicity when delivered as DNA. We detected no significant enhancement of RT-specific immune response to RT1.14oil compared to RT1.14opt-in. Both genes induced a similar loss in the expression of luciferase reporter from the sites of immunization indicating that an immune response induced by RT-14opt-in and by RT1.14oil genes had a similar capacity for in vivo clearance of RT-expressing cells. The only difference was observed in the profile of the secreted effector molecules. Mice DNA immunized with RT1.14oil responded by the secretion of perforin, in which the latter correlated to the levels of RT-specific dual secretion of IFN-γ/IL-2. In contrast, mice DNA immunized with RT1.14opt-in responded by the RT-specific secretion of granzyme B not correlated to the production of either IFN-γ or IL-2. Thus, the enhancement of secretion due to signal peptide tag led to no major changes in the profile of anti-RT cellular immune response with the exception of a shift from the secretion of granzyme B in response to RT1.14opt-in, in favor of perforin in the response to RT1.14oil.

The addition of secretion signal is expected to promote antibody formation. Targeting for secretion of the membrane-bound glycoprotein of viral haematopoietic septicaemia virus resulted in the enhancement of IgM response in the rainbow trout [57]. DNA immunization with the cytoplasmic protein cathepsin B fused to the signal peptide of TPA induced a twofold increase in the level of specific IgG [58]. However, here, secretion of RT1.14 caused no enhancement of either IgG or IgA responses. Similar observations were made earlier for the rabies glycoprotein that was originally exposed on the cell surface and elicited strong antibody response even prior to artificial secretion; addition of a secretion signal gave no boost to the specific antibody response [59]. Apparently, even comparatively low amounts of extracellular immunogen (as in case of RT1.14opt-in) are capable of eliciting high antibody production. Further increase in the amounts of extracellular protein due to artificial secretion do not promote further enhancement of humoral responses. Of note, signal sequence of TBEV NS1 appeared to be nonimmunogenic on either cellular or antibody levels, the latter favoring its inclusion into the heterologous protein immunogens as a noninterfering transport moiety.

Both the parental and the signal peptide-tagged RTs were secreted. Enhancement of secretion due to signal peptide tag induced only moderate decrease in the oxidative stress. Hereby, we could extend our statement on the capacity of RTs to induce oxidative stress: it is induced by a wide panel of RT variants, wild-type and drug-resistant, expressed from viral and from the expression-optimized genes [19], preferably secreted as well as preferably intracellular, both

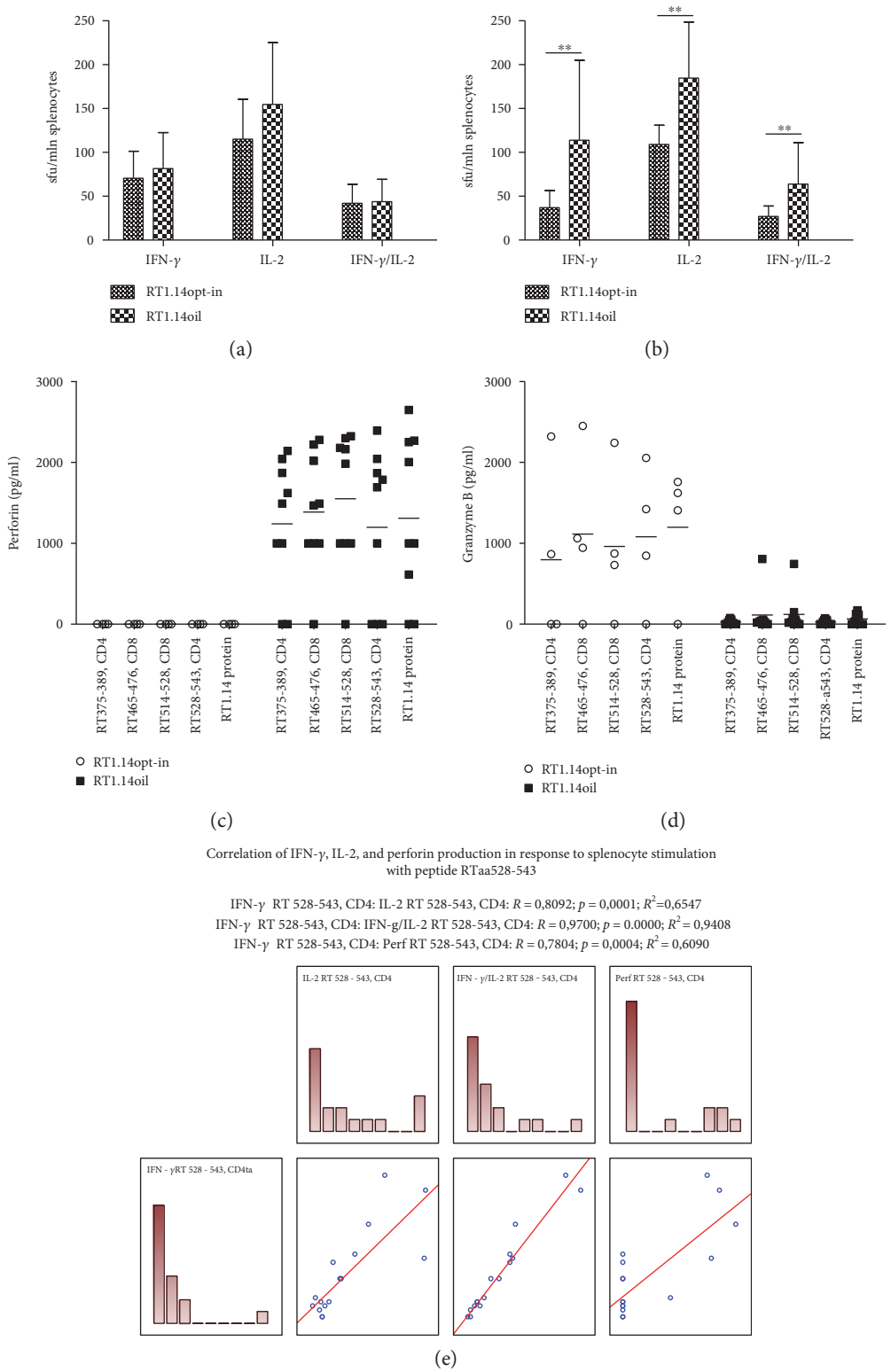


Figure 7: Cellular immune response of mice immunized with pVaxRT1.14opt-in and pVaxRT1.14oil. IFN-γ, IL-2, and dual (IFN-γ/IL-2) secretion by splenocytes in response to stimulation with RT1.14 protein (a) or RT1.14-derived peptide aa528-543 (b) measured by Fluorospot. Responses represent the average number of signal-forming units (sfu) per mln cells in two independent experiment runs, each done in duplicate, +SD, $n = 4$, $^{**}p < 0{,}1$ (Mann–Whitney test). Secretion of perforin (c) and granzyme B (d) by splenocytes of mice immunized with pVaxRT1.14opt-in and pVaxRT1.14oil in response to stimulation with RT1.14 protein and RT1.14-derived peptides. Splenocytes were stimulated with the recombinant RT 1.14 protein and RT peptides for 3 days. After that, cell culture fluids were collected and subjected to the analysis for granzyme B and perforin by sandwich ELISA. Data represent an average value of two repeated measurements for each mouse, in pg/ml. Correlation of the parameters of cellular immune response (e).

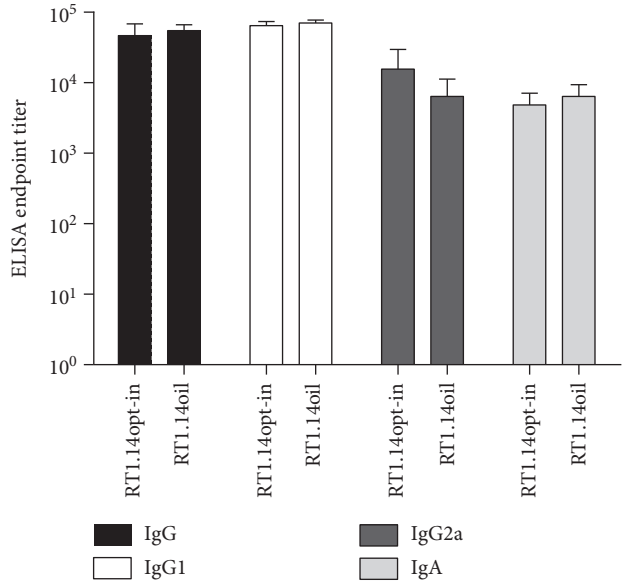


Figure 8: RT-specific IgGs and IgA induced in sera of mice immunized with pVaxRT1.14opt-in and pVaxRT1.14oil plasmids. Endpoint titers of RT-specific total IgG, IgG subtypes (IgG1, IgG2a), and IgA antibodies detected in the sera of BALB/c mice immunized with the genes encoding RT1.14opt-in and RT1.14oil plasmids. Data represent a mean + SD for the endpoint titer of antibodies against recombinant RT1.14 protein for four mice per group in two independent immunization runs. Cutoffs were set against the control mice immunized with empty vector pVax1 (see "Materials and Methods"). Groups demonstrate no significant difference in either IgG, or IgG1, or IgG2a, or IgA levels ($p > 0{,}05$).

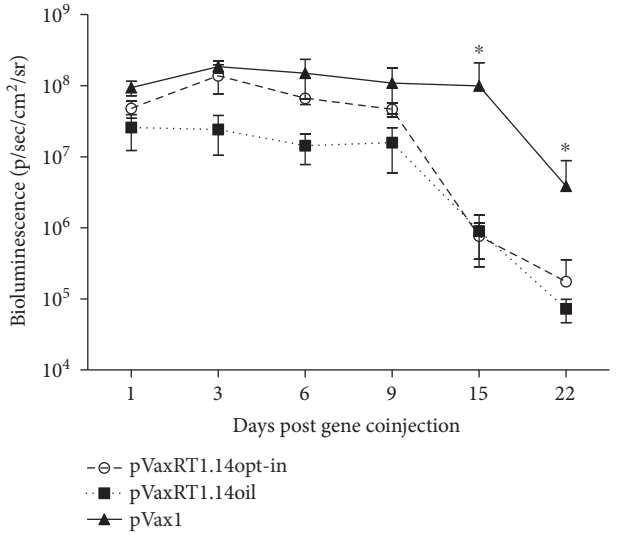


Figure 9: Kinetics of in vivo bioluminescence from the sites of coadministration of RT genes and luciferase reporter genes. In vivo monitoring of luciferase activity on days 1, 3, 6, 9, 15, 22 after the administration of plasmids encoding RT1.14opt-in and RT1.14oil, or empty vector pVax1, each mixed with pVaxLuc encoding firefly luciferase (1 : 1 *w*/*w*). One curve represents bioluminescent emission from eight mice with two immunization sites in each, assessed in two independent immunization runs. $^*p < 0{,}05$ (Mann–Whitney test).

enzymatically active and inactive. Altogether, this indicates that the propensity to induce oxidative stress and to modulate the immune response towards the antibody type is a property of the protein not linked to its enzymatic activities which can be modified and even abrogated.

We tentatively allocated this property to the RNase H domain of HIV-1 RT. RNases are necessary for the maintenance of cellular homeostasis, they participate in the degradation of ribosomes/cleavage of tRNA; RNases are involved in the recycling of phosphate during processes involving cell death and, hence, in the control of apoptosis [60]. Cleavage of tRNA by RNases is a conserved aspect of the response to oxidative stress [61, 62]. A secreted ribonuclease angiogenin selectively cleaves tRNA in eukaryotic cells; treatment of cells with this recombinant RNase or cell transfection with products of angiogenin activity (tRNA fragments) inhibits protein synthesis and induces apoptosis [62]. Overexpression of another RNase, RNase T2 family Rny1p, also caused apoptosis in yeast cells, whereas the deletion of its gene inhibited the apoptosis, specifically apoptosis in response to stress [63]. A proapoptotic activity of HIV-1 RNAse H was recently described [62]. Interestingly, Rny1p induced apoptosis independently of the catalytic activity through (yet unknown) interactions with the downstream components of the catabolic cell cascades [63, 64]. These observations indicate that oxidative stress induced by HIV RT expression may be linked to the properties of the RNase H moiety of the enzyme, not necessarily its activity. As Rny1p, inactive HIV-1 RNase H may retain the capacity to interact with the cellular proteins inducing oxidative stress and apoptotic cell death. By artificial secretion of HIV RT, we made it more stable and failed to reduce its content and, hence, the content of RNase H inside the cells, which in its turn did not allow us to significantly improve RT immunogenicity. Indirect confirmation of this concept lies in the experiments by Hallengard et al. who succeeded in raising strong immune response against RT after truncation of a part of RNase H domain [17]. Furthermore, we believe that the secretion that we found to be characteristic to this enzyme may play a role in the development of oxidative stress and oxidative stress response. Two other HIV proteins known to induce oxidative stress which contributes to HIV pathogenicity, namely, Tat and Vpr, are also secreted [65, 66]. Entry into the intercellular space turns Tat, Vpr, and possibly also RT into the signal molecules capable of affecting neighboring, also uninfected, cells. The actual mechanism(s) of the secretion of untagged RT and the input of this process into the induction of oxidative stress and modulation of immune response deserve a separate study.

Conflicts of Interest

The authors have no conflicts of interest to disclose.

Authors' Contributions

Maria Isaguliants and Elizaveta Starodubova shared last coauthorship.

Acknowledgments

This study was supported by the Russian Foundation of Basic Research, Grant no. 14-04-01817; immunization of mice and studies of anti-RT immune response were supported by the Russian Science Foundation, Grant no. 15-15-30039. Interaction of the partners and in-learning of the methods was supported by the EU Twinning Project VACTRAIN no. 692293 and the Swedish Institute PI Project no. 19806_2016.

References

[1] X. Jin, C. Morgan, X. Yu et al., "Multiple factors affect immunogenicity of DNA plasmid HIV vaccines in human clinical trials," *Vaccine*, vol. 33, no. 20, pp. 2347–2353, 2015.

[2] P. J. Munseri, A. Kroidl, C. Nilsson et al., "Priming with a simplified intradermal HIV-1 DNA vaccine regimen followed by boosting with recombinant HIV-1 MVA vaccine is safe and immunogenic: a phase IIa randomized clinical trial," *PLoS One*, vol. 10, no. 4, article e0119629, 2015.

[3] Y. Chen, S. Wang, and S. Lu, "DNA immunization for HIV vaccine development," *Vaccine*, vol. 2, no. 1, pp. 138–159, 2014.

[4] T. Ahmed, N. J. Borthwick, J. Gilmour, P. Hayes, L. Dorrell, and T. Hanke, "Control of HIV-1 replication in vitro by vaccine-induced human CD8(+) T cells through conserved subdominant pol epitopes," *Vaccine*, vol. 34, no. 9, pp. 1215–1224, 2016.

[5] N. Borthwick, T. Ahmed, B. Ondondo et al., "Vaccine-elicited human T cells recognizing conserved protein regions inhibit HIV-1," *Molecular Therapyy*, vol. 22, no. 2, pp. 464–475, 2014.

[6] A. Joachim, C. Nilsson, S. Aboud et al., "Potent functional antibody responses elicited by HIV-I DNA priming and boosting with heterologous HIV-1 recombinant MVA in healthy Tanzanian adults," *PloS One*, vol. 10, no. 4, article e0118486, 2015.

[7] D. Hallengard, B. K. Haller, S. Petersson et al., "Increased expression and immunogenicity of HIV-1 protease following inactivation of the enzymatic activity," *Vaccine*, vol. 29, no. 4, pp. 839–848, 2011.

[8] O. Krotova, E. Starodubova, S. Petkov et al., "Consensus HIV-1 FSU-A integrase gene variants electroporated into mice induce polyfunctional antigen-specific CD4+ and CD8+ T cells," *PloS One*, vol. 8, no. 5, article e62720, 2013.

[9] M. L. Knudsen, A. Mbewe-Mvula, M. Rosario et al., "Superior induction of T cell responses to conserved HIV-1 regions by electroporated alphavirus replicon DNA compared to that with conventional plasmid DNA vaccine," *Journal of Virology*, vol. 86, no. 8, pp. 4082–4090, 2012.

[10] A. Boberg, D. Sjostrand, E. Rollman, J. Hinkula, B. Zuber, and B. Wahren, "Immunological cross-reactivity against a drug mutated HIV-1 protease epitope after DNA multi-CTL epitope construct immunization," *Vaccine*, vol. 24, no. 21, pp. 4527–4530, 2006.

[11] E. Rollman, A. Brave, A. Boberg et al., "The rationale behind a vaccine based on multiple HIV antigens," *Microbes and Infection*, vol. 7, no. 14, pp. 1414–1423, 2005.

[12] E. S. Starodubova, A. Boberg, M. Litvina et al., "HIV-1 reverse transcriptase artificially targeted for proteasomal degradation induces a mixed Th1/Th2-type immune response," *Vaccine*, vol. 26, no. 40, pp. 5170–5176, 2008.

[13] E. Starodubova, A. Boberg, A. Ivanov et al., "Potent cross-reactive immune response against the wild-type and drug-resistant forms of HIV reverse transcriptase after the chimeric gene immunization," *Vaccine*, vol. 28, no. 8, pp. 1975–1986, 2010.

[14] T. J. Garrod, T. Gargett, W. Yu et al., "Loss of long term protection with the inclusion of HIV pol to a DNA vaccine encoding gag," *Virus Research*, vol. 192, pp. 25–33, 2014.

[15] A. Boberg and M. Isaguliants, "Vaccination against drug resistance in HIV infection," *Expert Review of Vaccines*, vol. 7, no. 1, pp. 131–145, 2008.

[16] M. G. Isaguliants, S. V. Belikov, E. S. Starodubova et al., "Mutations conferring drug resistance affect eukaryotic expression of HIV type 1 reverse transcriptase," *AIDS Research and Human Retroviruses*, vol. 20, no. 2, pp. 191–201, 2004.

[17] D. Hallengard, B. Wahren, and A. Brave, "A truncated plasmid-encoded HIV-1 reverse transcriptase displays strong immunogenicity," *Viral Immunology*, vol. 26, no. 2, pp. 163–166, 2013.

[18] M. G. Isaguliants, B. Zuber, A. Boberg et al., "Reverse transcriptase-based DNA vaccines against drug-resistant HIV-1 tested in a mouse model," *Vaccine*, vol. 22, no. 13-14, pp. 1810–1819, 2004.

[19] M. Isaguliants, O. Smirnova, A. V. Ivanov et al., "Oxidative stress induced by HIV-1 reverse transcriptase modulates the enzyme's performance in gene immunization," *Human Vaccines & Immunotherapeutics*, vol. 9, no. 10, pp. 2111–2119, 2013.

[20] A. V. Ivanov, V. T. Valuev-Elliston, O. N. Ivanova et al., "Oxidative stress during HIV infection: mechanisms and consequences," *Oxidative Medicine and Cellular Longevity*, vol. 2016, p. 8910396, 2016.

[21] G. Winkler, V. B. Randolph, G. R. Cleaves, T. E. Ryan, and V. Stollar, "Evidence that the mature form of the flavivirus nonstructural protein NS1 is a dimer," *Virology*, vol. 162, no. 1, pp. 187–196, 1988.

[22] J. M. Lee, A. J. Crooks, and J. R. Stephenson, "The synthesis and maturation of a non-structural extracellular antigen from tick-borne encephalitis virus and its relationship to the intracellular NS1 protein," *The Journal of General Virology*, vol. 70, Part 2, pp. 335–343, 1989.

[23] M. Flamand, F. Megret, M. Mathieu, J. Lepault, F. A. Rey, and V. Deubel, "Dengue virus type 1 nonstructural glycoprotein NS1 is secreted from mammalian cells as a soluble hexamer in a glycosylation-dependent fashion," *Journal of Virology*, vol. 73, no. 7, pp. 6104–6110, 1999.

[24] S. E. S. Kuzmenko Y. V., A. S. Shevtsova, G. G. Karganova, A. V. Timofeev, and V. L. Karpov, "Protectivity of gene-based constructions carrying a gene of NS 1 protein of tick-borne encephalitis virus," *BioDrugs*, vol. 1-2, no. 33-34, pp. 7–9, 2009.

[25] W. F. Fan and P. W. Mason, "Membrane association and secretion of the Japanese encephalitis virus NS1 protein from cells expressing NS1 cDNA," *Virology*, vol. 177, no. 2, pp. 470–476, 1990.

[26] B. Falgout, R. Chanock, and C. J. Lai, "Proper processing of dengue virus nonstructural glycoprotein NS1 requires the N-terminal hydrophobic signal sequence and the downstream nonstructural protein NS2a," *Journal of Virology*, vol. 63, no. 5, pp. 1852–1860, 1989.

[27] M. R. King, A. S. Ismail, L. S. Davis, and D. R. Karp, "Oxidative stress promotes polarization of human T cell differentiation

toward a T helper 2 phenotype," *Journal of Immunology*, vol. 176, no. 5, pp. 2765–2772, 2006.

[28] E. N. Wardle, *Guide to Signal Pathways in Immune Cells*, Humana Press, New York, 2009.

[29] M. Rath, I. Muller, P. Kropf, E. I. Closs, and M. Munder, "Metabolism via arginase or nitric oxide synthase: two competing arginine pathways in macrophages," *Frontiers in Immunology*, vol. 5, p. 532, 2014.

[30] A. V. Ivanov, A. N. Korovina, V. L. Tunitskaya et al., "Development of the system ensuring a high-level expression of hepatitis C virus nonstructural NS5B and NS5A proteins," *Protein Expression and Purification*, vol. 48, no. 1, pp. 14–23, 2006.

[31] S. F. Le Grice and F. Gruninger-Leitch, "Rapid purification of homodimer and heterodimer HIV-1 reverse transcriptase by metal chelate affinity chromatography," *European Journal of Biochemistry/FEBS*, vol. 187, no. 2, pp. 307–314, 1990.

[32] R. S. Fletcher, G. Holleschak, E. Nagy et al., "Single-step purification of recombinant wild-type and mutant HIV-1 reverse transcriptase," *Protein Expression and Purification*, vol. 7, no. 1, pp. 27–32, 1996.

[33] K. Rengan, "Cerenkov counting technique for beta particles: advantages and limitations," *Journal of Chemical Education*, vol. 60, no. 8, pp. 682–684, 1983.

[34] M. G. Isaguliants, K. Pokrovskaya, V. I. Kashuba et al., "Eukaryotic expression of enzymatically active human immunodeficiency virus type 1 reverse transcriptase," *FEBS Letters*, vol. 447, no. 2-3, pp. 232–236, 1999.

[35] A. S. Rytting, L. Akerblom, J. Albert et al., "Monoclonal antibodies to native HIV type 1 reverse transcriptase and their interaction with enzymes from different subtypes," *AIDS Research and Human Retroviruses*, vol. 16, no. 13, pp. 1281–1294, 2000.

[36] A. V. Ivanov, O. A. Smirnova, O. N. Ivanova, O. V. Masalova, S. N. Kochetkov, and M. G. Isaguliants, "Hepatitis C virus proteins activate NRF2/ARE pathway by distinct ROS-dependent and independent mechanisms in HUH7 cells," *PloS One*, vol. 6, no. 9, article e24957, 2011.

[37] P. Zhou, "Determining protein half-lives," *Methods in Molecular Biology*, vol. 284, pp. 67–77, 2004.

[38] A. K. Roos, F. Eriksson, D. C. Walters, P. Pisa, and A. D. King, "Optimization of skin electroporation in mice to increase tolerability of DNA vaccine delivery to patients," *Molecular Therapy*, vol. 17, no. 9, pp. 1637–1642, 2009.

[39] J. K. Wakefield, S. A. Jablonski, and C. D. Morrow, "In vitro enzymatic activity of human immunodeficiency virus type 1 reverse transcriptase mutants in the highly conserved YMDD amino acid motif correlates with the infectious potential of the proviral genome," *Journal of Virology*, vol. 66, no. 11, pp. 6806–6812, 1992.

[40] G. L. Beilhartz and M. Gotte, "HIV-1 ribonuclease H: structure, catalytic mechanism and inhibitors," *Virus*, vol. 2, no. 4, pp. 900–926, 2010.

[41] K. R. Young, J. M. Smith, and T. M. Ross, "Characterization of a DNA vaccine expressing a human immunodeficiency virus-like particle," *Virology*, vol. 327, no. 2, pp. 262–272, 2004.

[42] J. M. Smith, R. R. Amara, H. M. McClure et al., "Multiprotein HIV type 1 clade B DNA/MVA vaccine: construction, safety, and immunogenicity in macaques," *AIDS Research and Human Retroviruses*, vol. 20, no. 6, pp. 654–665, 2004.

[43] L. A. Kohlstaedt, J. Wang, J. M. Friedman, P. A. Rice, and T. A. Steitz, "Crystal structure at 3.5 a resolution of HIV-1 reverse transcriptase complexed with an inhibitor," *Science*, vol. 256, no. 5065, pp. 1783–1790, 1992.

[44] Z. Hostomska, D. A. Matthews, J. F. Davies 2nd, B. R. Nodes, and Z. Hostomsky, "Proteolytic release and crystallization of the RNase H domain of human immunodeficiency virus type 1 reverse transcriptase," *The Journal of Biological Chemistry*, vol. 266, no. 22, pp. 14697–14702, 1991.

[45] G. Haas, R. David, R. Frank et al., "Identification of a major human immunodeficiency virus-1 reverse transcriptase epitope recognized by mouse CD4+ T lymphocytes," *European Journal of Immunology*, vol. 21, no. 6, pp. 1371–1377, 1991.

[46] C. M. Snapper and W. E. Paul, "Interferon-gamma and B cell stimulatory factor-1 reciprocally regulate Ig isotype production," *Science*, vol. 236, no. 4804, pp. 944–947, 1987.

[47] E. Starodubova, O. Krotova, D. Hallengard et al., "Cellular immunogenicity of novel gene immunogens in mice monitored by in vivo imaging," *Molecular Imaging*, vol. 11, no. 6, pp. 471–486, 2012.

[48] S. Sutti, A. Jindal, I. Locatelli et al., "Adaptive immune responses triggered by oxidative stress contribute to hepatic inflammation in NASH," *Hepatology*, vol. 59, no. 3, pp. 886–897, 2014.

[49] M. Los, W. Droge, K. Stricker, P. A. Baeuerle, and K. Schulze-Osthoff, "Hydrogen peroxide as a potent activator of T lymphocyte functions," *European Journal of Immunology*, vol. 25, no. 1, pp. 159–165, 1995.

[50] A. Csillag, I. Boldogh, K. Pazmandi et al., "Pollen-induced oxidative stress influences both innate and adaptive immune responses via altering dendritic cell functions," *Journal of Immunology*, vol. 184, no. 5, pp. 2377–2385, 2010.

[51] A. Larbi, F. Cabreiro, H. Zelba et al., "Reduced oxygen tension results in reduced human T cell proliferation and increased intracellular oxidative damage and susceptibility to apoptosis upon activation," *Free Radical Biology & Medicine*, vol. 48, no. 1, pp. 26–34, 2010.

[52] D. M. Lowe, V. Parmar, S. D. Kemp, and B. A. Larder, "Mutational analysis of two conserved sequence motifs in HIV-1 reverse transcriptase," *FEBS Letters*, vol. 282, no. 2, pp. 231–234, 1991.

[53] C. C. King, M. M. Jefferson, and E. L. Thomas, "Secretion and inactivation of myeloperoxidase by isolated neutrophils," *Journal of Leukocyte Biology*, vol. 61, no. 3, pp. 293–302, 1997.

[54] L. M. Costantini, M. Baloban, M. L. Markwardt et al., "A palette of fluorescent proteins optimized for diverse cellular environments," *Nature Communications*, vol. 6, p. 7670, 2015.

[55] J. Tyedmers, A. Mogk, and B. Bukau, "Cellular strategies for controlling protein aggregation," *Nature Reviews. Molecular Cell Biology*, vol. 11, no. 11, pp. 777–788, 2010.

[56] A. Ganesan, A. Siekierska, J. Beerten et al., "Structural hot spots for the solubility of globular proteins," *Nature Communications*, vol. 7, p. 10816, 2016.

[57] A. Martinez-Lopez, P. Garcia-Valtanen, M. Ortega-Villaizan Mdel et al., "Increasing versatility of the DNA vaccines through modification of the subcellular location of plasmid-encoded antigen expression in the in vivo transfected cells," *PloS One*, vol. 8, no. 10, article e77426, 2013.

[58] R. Jayaraj, D. Piedrafita, T. Spithill, and P. Smooker, "Evaluation of the immune responses induced by four targeted DNA vaccines encoding the juvenile liver fluke antigen, cathepsin B in a mouse model," *Genetic Vaccines and Therapy*, vol. 10, no. 1, p. 7, 2012.

[59] M. Kaur, A. Rai, and R. Bhatnagar, "Rabies DNA vaccine: no impact of MHC class I and class II targeting sequences on immune response and protection against lethal challenge," *Vaccine*, vol. 27, no. 15, pp. 2128–2137, 2009.

[60] G. C. MacIntosh, *RNase T2 Family: Enzymatic Properties, Functional Diversity, and Evolution of Ancient Ribonucleases*, vol. 26, Springer-Verlag, Berlin Heidelberg, 2011.

[61] D. M. Thompson, C. Lu, P. J. Green, and R. Parker, "tRNA cleavage is a conserved response to oxidative stress in eukaryotes," *RNA*, vol. 14, no. 10, pp. 2095–2103, 2008.

[62] S. Yamasaki, P. Ivanov, G. F. Hu, and P. Anderson, "Angiogenin cleaves tRNA and promotes stress-induced translational repression," *The Journal of Cell Biology*, vol. 185, no. 1, pp. 35–42, 2009.

[63] Y. B. Ouyang, S. G. Carriedo, and R. G. Giffard, "Effect of Bcl-x(L) overexpression on reactive oxygen species, intracellular calcium, and mitochondrial membrane potential following injury in astrocytes," *Free Radical Biology & Medicine*, vol. 33, no. 4, pp. 544–551, 2002.

[64] L. B. Ray, "RNases signal stress," *Science Signaling*, vol. 2, no. 66, p. ec134, 2009.

[65] K. M. Porter and R. L. Sutliff, "HIV-1, reactive oxygen species, and vascular complications," *Free Radical Biology & Medicine*, vol. 53, no. 1, pp. 143–159, 2012.

[66] M. R. N. Adriano Ferrucci and B. Wigdahl, "Human immunodeficiency virus viral protein R as an extracellular protein in neuropathogenesis," *Advances in Virus Research*, A. J. S. Karl Maramorosch and F. A. Murphy, Eds., vol. 81, pp. 165–199, 2011.

Hindawi
Journal of Immunology Research
Volume 2017, Article ID 5974574, 13 pages
https://doi.org/10.1155/2017/5974574

Review Article

The Immune Epitope Database: How Data Are Entered and Retrieved

Ward Fleri, Kerrie Vaughan, Nima Salimi, Randi Vita, Bjoern Peters, and Alessandro Sette

Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, 9420 Athena Circle, La Jolla, CA 92037, USA

Correspondence should be addressed to Ward Fleri; wfleri@lji.org

Received 28 February 2017; Accepted 26 April 2017; Published 29 May 2017

Academic Editor: Pedro A. Reche

Easy access to a vast collection of experimental data on immune epitopes can greatly facilitate the development of therapeutics and vaccines. The Immune Epitope Database and Analysis Resource (IEDB) was developed to provide such a resource as a free service to the biomedical research community. The IEDB contains epitope and assay information related to infectious diseases, autoimmune diseases, allergic diseases, and transplant/alloantigens for humans, nonhuman primates, mice, and any other species studied. It contains T cell, B cell, MHC binding, and MHC ligand elution experiments. Its data are curated primarily from the published literature and also include direct submissions from researchers involved in epitope discovery. This article describes the process of capturing data from these sources and how the information is organized in the IEDB data. Different approaches for querying the data are then presented, using the home page search interface and the various specialized search interfaces. Specific examples covering diverse applications of interest are given to highlight the power and functionality of the IEDB.

1. A High-Level Overview of the IEDB

The Immune Epitope Database (IEDB) is a free online resource that catalogs and makes accessible to the scientific community epitope-related data derived from allergic diseases, infectious diseases, apart from HIV which is captured separately in the Los Alamos HIV database [1], autoimmune diseases, and diseases associated with transplantation and alloantigens. The IEDB [2] contains both T cell and B cell epitopes, as well as MHC ligand data. The epitopes are derived from humans, nonhuman primates, mice, and all other studied hosts.

The curation of scientific literature started in 2004, requiring the curation of past and current relevant epitope literature in available peer-reviewed journals [3, 4]. As the IEDB evolved, it has been necessary to change how biological concepts are captured in order to maximize accuracy. In addition, automated validation is continuously added. Consequently, there is a significant ongoing "recuration" effort of revising existing entries to improve data quality and consistency. The IEDB is now current with the published literature, and targeted PubMed queries are run biweekly, with the goal of making the data available in the IEDB within eight weeks of publication.

As of November 2016, the IEDB data were derived from over 18,000 papers. In addition, the IEDB contains nearly 300 submissions (corresponding to approximately 20% of the total data) from several NIH-funded large-scale epitope discovery programs and from researchers that directly approach the IEDB to deposit their data, including negative data, which might typically appear in supplemental tables or might not be published at all. Slightly more than half of the references in the IEDB relate to infectious diseases and about a quarter relate to autoimmune diseases. The remainder includes allergy, transplant, and other categories. Because the data in the IEDB reside in the public domain, researchers can freely access, analyze, and publish works using these data.

When initially designing the IEDB, we realized that different researchers had different views on what should be included in an epitope database, based largely on their area of research. While each scientist's specific interests and epitope definitions might vary, ultimately every epitope is defined by an experiment or specific assay. To provide a general yet accurate epitope database, we utilized an assay-

centric design, capturing the experiments (assays) that characterize and define each epitope. This required translation of the data typically described in the methods and results sections of a scientific paper into a generic data structure in which epitope data are entered and stored in a format that allows users to query for the characteristics of these epitopes.

The IEDB website has undergone two major revisions since its introduction in early 2006, each improving its features and usability [2, 5]. The current IEDB 3.0, deployed in February 2015, incorporated feedback from immunologists and bioinformaticians, collected at user workshops, help desk requests, and user observation sessions, to make searching the database more intuitive and to deliver results in a more useful format. The home page prominently features the query interface, which contains the fields that address over 80% of the typical queries. The query concept is similar to a travel website where users specify basic information and can then filter or refine the results with additional fields. In the following sections, we will present a detailed account of the IEDB query and reporting and also provide several specific examples.

2. Which Data Are Included in the IEDB?

The IEDB does not capture predicted/inferred data or reviews. All data come from empirically derived assays reflecting the interaction of an adaptive immune receptor with an epitope. There are four categories of assays—T cell, B cell (including soluble antibodies), MHC binding, and MHC ligand elution.

Data are captured in accordance with clearly established criteria [5]. Linear peptides cannot exceed 50 amino acids in length and must be tested as either an immunogen or an antigen. We capture discontinuous epitopes where they show that specific residues are important for antibody recognition of an antigen, usually through mutation, or from crystal structures of B cell receptor and peptide. The database also includes nonpeptidic epitopes. These include carbohydrates, lipids, chemicals, metals, and drugs. In addition, for all data there is a minimal set of information required. For linear peptides, the amino acid sequence must be specified, and for discontinuous epitopes, the amino acid and their respective position must be specified (e.g., G104, G106, L107, and W231). In all cases, it must be possible to determine whether the outcome was positive or negative. In the case of T and B cell assays, a host (e.g., human or mouse) must be specified.

Because the IEDB is fundamentally a database of experiments, each paper or reference to be included must describe at least one epitope, and each epitope must be described in one or more experiments. Up to 400 data fields are used to capture all the details of these experiments. Simplifying the curation process and ensuring accuracy and consistency is largely addressed by the use of external resources and ontologies [6].

The goal of the IEDB is to allow a global query of all data from the scientific literature and direct submission. Individual epitopes can be tested in multiple references, in more than one host and under different conditions. If an epitope has been studied in multiple publications, users can gather all the cumulative data from the IEDB and draw their own conclusions regarding its utility for their application. The same epitope might be tested in different assay types or may be repeatedly tested in the same experimental setup by different groups. In some cases, the assay may be positive and in others negative and the user can examine the experimental details captured therein to critically evaluate the available information.

3. How Are the Data Structured?

As stated above, each IEDB record must be associated with a clearly defined molecular structure. In addition, for peptidic epitopes, the author must provide a sequence and identify its protein source and organism source. While some authors specify a GenBank or UniProt identifier for each protein, most provide a protein name or abbreviation. Since protein nomenclature is highly variable and a search for protein names in GenBank can yield a large number of results, the protein source assigned to a given peptide by the IEDB must have a 100% BLAST match [7] with that sequence and must correspond to the protein that the author indicated.

Each peptidic epitope is also derived from an organism species, and authors may further specify a particular strain. The IEDB uses the NCBI taxonomy (https://www.ncbi.nlm.nih.gov/taxonomy) to formalize the organism assignment. The extensive taxonomy tree has been modified to only include species found in the IEDB and subsequently embedded in an organism finder, augmented with synonyms and auto-complete features to facilitate queries.

The curation of nonpeptidic structures within the IEDB conforms with the Chemical Entities of Biological Interest (ChEBI) database in the UK (https://www.ebi.ac.uk/chebi/) [8]. For this, a ChEBI curator with specific chemistry expertise assists in describing each chemical structure captured in the database. In the case of nonpeptidic epitopes, the ChEBI curator reads the paper, draws the structure, deposits it into the ChEBI database, and provides the ChEBI identifier, which is then used to curate all associated assays with that structure [9].

To describe the immunization processes, the NCBI taxonomy tree is again used to specify the host organism whose T cells or antibodies are being studied in experiment. The immunization process describes how the host became initially exposed or sensitized. For example, a mouse was injected with a protein, a human was given a vaccine, or a human contracted a specific disease. The Disease Ontology (DO) [10] is used to provide a standardized list of disease names and identifiers, which enables linkage of the IEDB data to other databases using the same identifiers. The use of external ontologies like DO provides useful information, such as definitions, synonyms, and relationships that the IEDB can leverage to enhance its information content.

The Ontology of Biomedical Investigation (OBI) [11], which gives a hierarchical tree structure of all assay types found in the literature, is used to capture the details of the experiment accurately and as completely as possible. The hierarchical tree structure allows users to search at many levels, for example, on all T cell assays, or specifically

cytokine assays, or only IL-2 assays. As we encounter new assay types, we request OBI to add new terms to the ontology, which can then be used by our curators and end users.

MHC restriction is an important parameter related to T cell assays, as well as MHC binding and ligand elution assays. Because no MHC allele ontology existed, the IEDB created the MHC Restriction Ontology (MRO). MRO models the protein complex of the MHC as being composed of two different protein chains that are each encoded by different loci that come together to form a complex [12]. Users can search by locus, protein chain, or the whole MHC complex. In addition, haplotype information is included for mice.

4. The Curation Process

The literature curation process starts with the execution of a complex keyword query of PubMed to retrieve potentially relevant papers. An automated classifier [13–15] then assesses the likelihood that a paper contains relevant information and assigns each paper retrieved to a disease category. Curatability is then reviewed by a staff immunologist, and papers are assigned to the curation team. Individual papers are read, assessed for meeting strict inclusion criteria (http://curationwiki.iedb.org/wiki/index.php/IEDB_Inclusion_Criteria), and then entered into the database. About 30% of the papers do not meet the inclusion criteria and are marked therefore as uncuratable. Although the abstract may contain the requisite keywords, the actual article commonly might omit sequence information or might not actually be about epitopes. Articles about NK epitopes, epitope tags, superantigens, and antigen processing are also excluded.

Curators use a specialized internal application called a "finder" to assign a GenBank (https://www.ncbi.nlm.nih.gov/genbank/) or UniProt (http://www.uniprot.org/) identifier to each epitope based on 100% sequence identity. Finders are incorporated into the curation system and the external production system to facilitate input selections and the use of controlled vocabulary. They organize various query parameters, such as molecules and organisms, in a hierarchical tree fashion. As an example, the hemagglutinin protein from influenza A virus has more than 300 different GenBank entries that capture differences in the epitope sequences encountered in the literature, but these are presented by the finder as a single hemagglutinin entry.

To avoid sorting through this many possible sequences of hemagglutinin when performing a query, the protein source data are organized using a UniProt reference proteome. UniProt provides a list of all the proteins for a given species that is used to group all the proteins used as epitope sources. Therefore, only the UniProt hemagglutinin is displayed in the finder (Figure 1). Stars (**) are used to indicate the quality and completeness of the associated reference proteome for that organism. Within the finder application, a question mark icon provides a link that elaborates on the star grading system. UniProt also provides information on the processing of proteins into functional fragments. The IEDB search interface embeds these UniProt annotations to enable searches on processed fragments or on full-length proteins. Users can also query all influenza virus proteins by selecting a higher node in the finder tree. Similar finders with hierarchical trees are used throughout the query interface.

All curated papers undergo peer review to determine if the captured data are accurate. If necessary, authors are contacted for clarifications on the paper's content. New data are promoted on the IEDB website weekly, representing 15–20 new papers per week.

The papers themselves can differ greatly, given the variety of journals, authors, and diseases, making data consistency a challenge. To address these differences, a number of quality control measures have been implemented. Formal curation guidelines have been developed over the years as we encounter new concepts and techniques in the literature (http://curationwiki.iedb.org/wiki/index.php/Curation_Manual2.0) [16]. We also refer to immunological experts as needed, especially when modifications of how data are captured are proposed or new assays or immunological content is encountered. To assist and streamline the curation process, a web-based curation system has been developed with built-in validation logic. For example, when a curator inputs a certain value in the data field, input for other data fields are constrained to certain allowable values.

5. Direct Submission Processes

As mentioned above, the IEDB is populated with data from direct submissions in addition to literature curation. Direct submissions account for about 20% of the epitope records. The primary source of submissions has been the NIAID-funded large-scale T cell and antibody epitope discovery contracts and grants for infectious diseases, such as dengue virus, influenza virus, arenavirus, tuberculosis, and common allergens, although other researchers have deposited their epitopes as well.

There are three ways to submit data to the IEDB. The first is a wizard system that walks the submitter through a step-by-step process to enter data. This is ideal for new submitters who need to become familiar with the data fields and for those with only a few epitopes and assays. The second method utilizes Excel files that work as templates for different types of assays. This method accommodates larger data submissions and is the most commonly used. The third method uses XML files, which offer the greatest flexibility, but also require expertise in the use of XML and familiarity with the IEDB data structure. One of the IEDB curators specializes in assisting researchers in submitting data.

Just as all curated published literature has a PubMed ID that is captured in the IEDB, all data submissions are assigned a submission ID that can be used in publications and later referenced. The submissions are reviewed by curators to ensure completeness and accuracy, and the data validation rules are executed as done for data curated from literature. The data are not made publicly available until the author requests it, usually at the time that a corresponding manuscript is published. If a data submission is associated with a publication curated by the IEDB, the submission ID is linked to the PubMed ID.

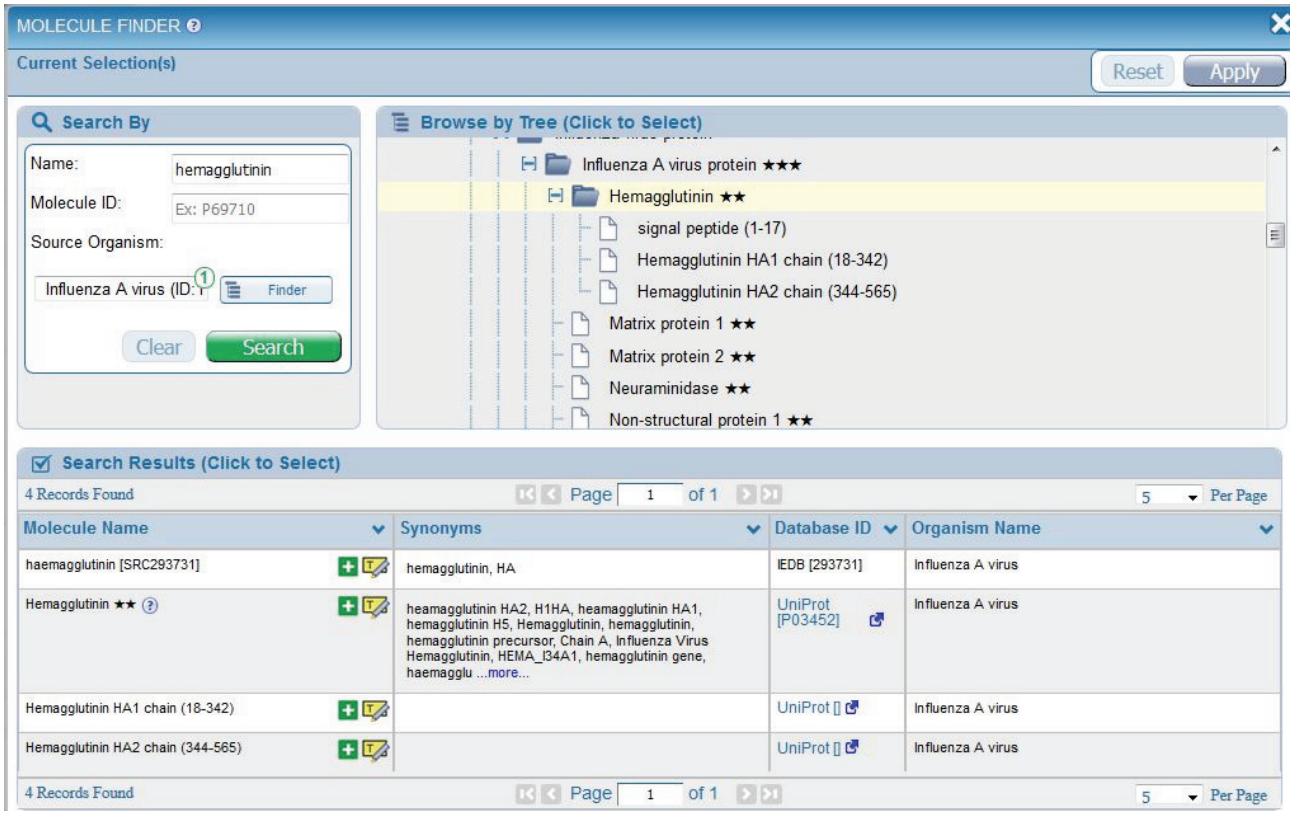


Figure 1: The Molecule Finder can be used to select the source antigen of the epitope. The yellow highlight icon in the bottom table highlights the selection in the tree in the upper right. The green "+" icon selects the molecule for the search criteria. The stars next to the molecule name indicate the quality and completeness of the associated reference proteome.

6. IEDB Search Basics

The IEDB provides users with two basic options—the home page search and specialized searches. In both cases, special "finder" features assist in this process. Data field names are explicit and concise by necessity and may be unfamiliar to some users. For example, it is necessary to distinguish whether a protein is delivered as a vaccine or if it is used on an ELISA plate. The protein used to immunize is called the immunogen, and the protein on the plate is called the antigen. The source of peptides is the epitope's source antigen. We also capture in vivo and in vitro processes. The former includes natural infection, administration of a vaccine, and occurrence of disease, while the latter includes T cell restimulation.

The search interface on the IEDB home page (Figure 2; http://www.iedb.org) allows performing the majority of user queries and provides the ability to refine initial queries on the subsequent results page. The home page interface contains six sections or panes that allow users to search by major categories of epitope, antigen, host, assay, MHC restriction, and disease. Clicking the green Search button brings the user to the results page. This page contains panels on the left-hand side that allow for further filtering of the results using the same six elements as before, plus a pane for references. Although some of these panes are similar to those on the home page, they allow users to be more specific in selecting search criteria and thus refining the query.

The query results appear on four different tabs. The Epitope tab lists the epitope description or sequence, as well as the antigen and organism from which it is derived. The Antigen tab contains the antigen name and corresponding organism. The Assay tab summarizes information for the T cell, B cell, and MHC ligand responses and respective journal articles. The Reference tab lists all the papers and submissions providing results for the query. The number of records in each tab is listed in parentheses. Users can sort results on these tabs by clicking on the column header. The results of each tab can be downloaded as comma-separated value (CSV) files by clicking on the Excel icon. They can then be opened by a spreadsheet program for further analysis and data manipulation.

All queries are run with positive assay results as the default because most people expect this behavior. However, it is worth stressing that the IEDB also captures negative experimental results. By making use of this information, researchers can decide not to repeat experiments that others have conducted. These data points are also important in training machine learning algorithms.

To simplify and at the same time ensure rigor in definitions and data searches, behind the scene, we make extensive use of ontological terminology, such as "parent-child" relationships. For example, the protein hemagglutinin has different hemagglutinin children that represent that protein for specific strains. The relationships between the epitope,

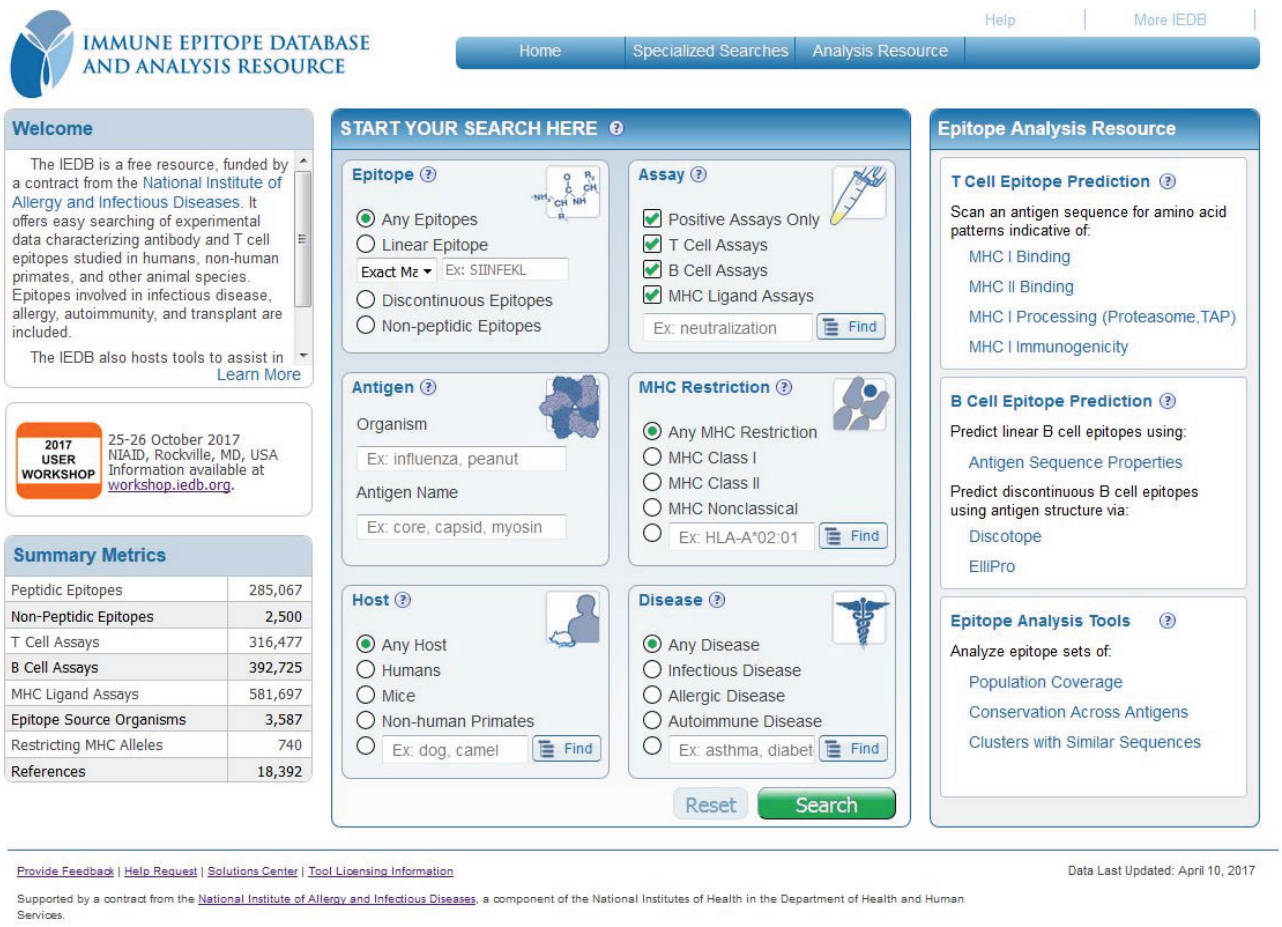


Figure 2: The IEDB home page features a search interface in the center of the page.

the protein from which it is derived, and the organism that is the source of the protein are all captured. Further description of the terms and nomenclature can be found on the Data Field Descriptions page (http://curationwiki.iedb.org/wiki/index.php/Data_Field_Descriptions) and in the IEDB Curation Manual (http://curationwiki.iedb.org/wiki/index.php/Curation_Manual2.0).

7. Easy Options to Narrow Search Results

As a user moves the mouse over each search pane on the results page, it expands to reveal free-text fields and finder buttons. In the Epitope pane, by default the radio button for "Any Epitopes" is selected, but users can restrict their search to linear, discontinuous, or nonpeptidic epitopes. By expanding this pane, users can specify a sequence for a linear epitope. They can also select whether they want an exact match to the sequence, find it as a substring in another epitope, or find epitopes that are homologous to it at 70%, 80%, or 90% sequence identity. A specific finder enables the selection of a nonpeptidic structure.

The Antigen pane has two fields, Organism and Antigen Name, which permit users to search for pathogen, allergen, or self-antigen from which the epitope was derived. For autoimmune diseases, these would be human proteins, and for allergens, these would include sources such as grass, pollen, or dust mites. The text fields are auto-complete enabled, so as the user starts typing, possible values will be listed below to help speed the input process. Synonyms are included, so typing HCV, for example, will produce Hepatitis C virus as an option.

The Assay pane by default selects all assay types that have positive measurements. The search can be refined by deselecting these options and by specifying assay types with the auto-complete text fields or finders for the three major assay types—T cell, B cell, and MHC ligand (binding and elution).

The MHC Restriction pane enables queries for the restrictions captured for a given epitope in a given T cell, MHC binding, or elution assay. Users can constrain the search at a high level to class I, class II, or nonclassical broad categories. They can also specify one or more alleles using the auto-complete text field or the finder.

The Host pane lets users define the host in which the epitope was described. The Host pane on the home page has selections for Any, Humans, Mice, Non-human Primates, and a text field for specifying other curated host species. On the results page, the Host pane replaces the Other Common Hosts with a Specific Host option with an accompanying auto-complete field and finder. This allows the specification of transgenic mice or particular mouse strains.

The Disease pane enables searching for data based on the clinical status of the host as detailed in the patient history or for a known animal model disease, such as experimental allergic encephalomyelitis (EAE) for multiple sclerosis. As such, it makes searching for autoimmune and allergens easier since antigens in these disease categories are not necessarily pathogenomonic, as they are with infectious disease antigens. Users can specify diseases with the auto-complete text field or the associated finder. Finally, the Reference pane can be used to limit searches to journal articles or submissions, if desired, or by author, title, year, or PubMed ID. In general, users can also specify author, reference titles, and year.

8. Using the Antigen, Organism, MHC Allele, Assay, and Disease Finders

As described above, finders provide standardization and hierarchical organization of the data in the IEDB and help the users formulate search criteria at different levels of granularity. The tree structure graphically displays parent-child relationships of the data. Finders exist for antigens, nonpeptidic molecules, organisms, MHC alleles, assays, and disease.

The Epitope pane contains a finder for nonprotein chemical entities, including metal allergens, such as nickel; drug compounds, such as penicillin; and glycolipids, such as lipopolysaccharides. The Antigen pane also contains an antigen finder for specifying the molecular source of the epitope (in most cases a protein). For example, a user interested in hemagglutinin for influenza A virus can type "HA" in the Name field and select "influenza A" using the auto-complete feature in the Source Organism text field and then click Search to generate a list of potential hemagglutinin hits.

As mentioned above, the organism finder is used to specify the source of the epitope or the assay immunogen or antigen. To do this, users can open the finder and type "HCV" in the Name text field in the upper left corner and click the Search button on the finder. This generates a list of related terms below. Users can then click on the green "+" icon to make a choice, which then appears at the top of the finder in Current Selections at top. Alternatively, a user can click on the yellow highlight icon, which will highlight the selected organism in the tree structure in the center (Figure 3). This highlight feature enables the user to visualize the taxonomic relationships and allows the user to select a strain, species, genus, or family. In either case, the user must click the green Apply button to complete the organism specification for that query. The organism finder can be accessed by the Antigen pane.

The Assay pane contains the assay finder that is used to find specific assay types within the broader category of B cell, T cell, or MHC ligand. Users can search by name, method, or "measurement of." In this way, users can specify assay methods such as ELISA, ELISPOT, ICS, or X-ray crystallography or can alternatively specify all assays that measure interferon gamma (IFNγ).

The allele finder in the Allele pane is used to specify the MHC restriction defined in the assay. As of January 2017, the IEDB MHC tree contains alleles from 16 different species, including primarily human, nonhuman primates, and mice. Users can type in the name of their alleles of interest in the Name text field or use the pull-down menus for Organism or Class and then click the Search button to generate a list of possible alleles. As with the other finders, users can select from the list or highlight an allele to expand the allele tree. The use of the finder is especially convenient since the allele notation can change over time.

The host finder on the Host pane allows users to select additional host species that are not available from the Host pane on the home page. This might include specific mouse strains, such as transgenic mouse species with complex names.

As discussed above, the disease finder within the Disease pane lets the user specify a disease, which is helpful when searching for autoimmune diseases or allergens. For example, a search for "asthma" using the Disease Name text field finds two types of asthma, allergic asthma and nonallergic/occupational asthma. Again, it should be emphasized that this field does not search for epitope data based on disease category but rather by the clinical state of the subject or animal model.

9. Specialized Search Options

While the home page search interface can accommodate the vast majority of common searches, more intricate or complex immunological queries may need one of the specialized search interfaces. The specialized searches can be accessed from the home page under the Specialized Searches pull-down menu. The first four items in the list, Epitope Details, T Cell Assay Details, B Cell Assay Details, and MHC Assay Details, allow the user to access every field in the database, greatly expanding the search options.

The first Specialized Search in the menu is Epitope Details. This page has a similar layout to the Results page in the home page search interface. It has the same four results tabs, but the filter selections on the left are different and can be expanded to reveal numerous fields that were not previously available. The Epitope Reference Details section contains fields for epitope name and start/stop positions in a reference protein. The Epitope Related Object section includes fields intended to capture entities that bear some relationship to the epitope—analog, mimotope, and neoepitope. Analogs are synthetic constructs of peptide sequences or chemical compounds that share some structural features in common with another sequence or compound. They are often used to determine the role of specific amino acids in the binding or immunogenicity of an epitope. The source of an analog is always artificial. Mimotopes are functional mimics of natural molecular structures which bear little or no sequence homology to their biological counterparts. Cancer neoepitopes are defined herein as any epitope comprising amino acids that are not encoded in the germ line genome but that arise due to somatic (nonsynonomous) mutations, excluding somatic rearrangements of TCR and Ig genes. The IEDB will curate both immunogenic neoepitopes and putative epitopes tested for recognition that were not found to be immunogenic. Users can search by author, keywords in abstracts, journal name, MeSH terms, year, and more.

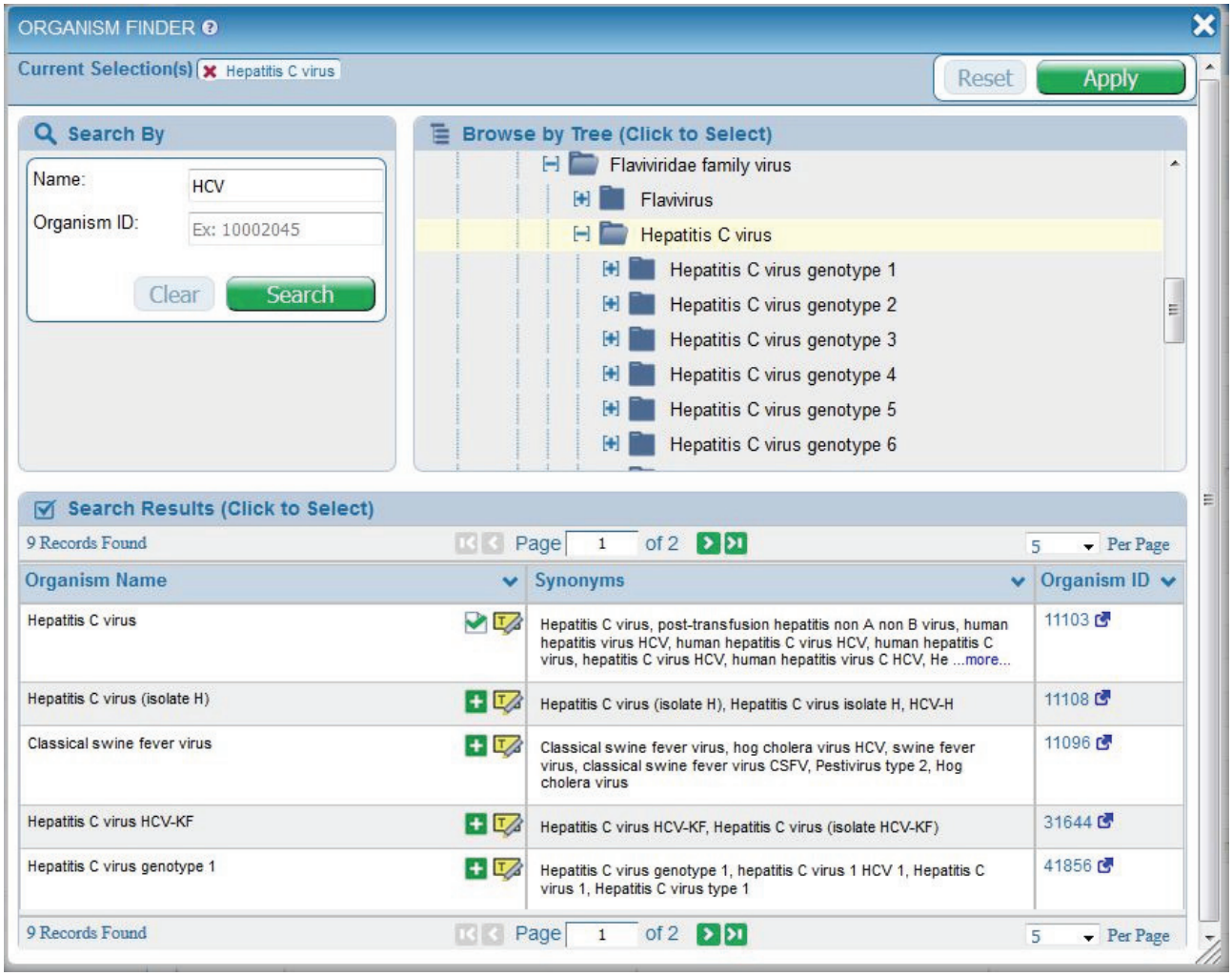


Figure 3: The Organism Finder can be used to select the search criteria for the epitope source organism. A similar finder is available for selecting the host organism.

The T Cell, B Cell, and MHC assay details pages all have the same capabilities described above plus two others—Host and Assay. The Host pane contains fields related to how the host acquired immune reactivity, such as being immunized or by natural infection. Users can also specify the immunogen (e.g., epitope, protein, or organism) and the host disease, among many other criteria. In the Assay pane, in addition to selecting assay type, users can specify effector cells, antibody isotypes, monoclonal or polyclonal response, and the assay antigen (e.g., epitope, protein, or organism).

The Host pane also has data fields for the first and second in vivo processes. An in vivo process indicates how the host acquired immune reactivity, such as by immunization or natural exposure. A host can also encounter an immunogen more than once, which is why two in vivo processes are available. For example, a mouse can be immunized with a peptide and then later challenged with an organism and the cells taken from the organism and assayed in vitro. That scenario is captured with two in vivo processes, the first for the immunization of the mouse and the second for the challenge. There are also fields to capture in vitro process because often there is a restimulation performed in vitro prior to the assay.

To specify in vivo processes, the Host pane contains a multiselect pull-down menu with a variety of choices that can be categorized into groups related to administration, occurrence, exposure, transplant/transfusion, no immunization, and unknown. Definitions for all the choices are provided in the IEDB Curation Manual. The 1st Immunogen section has fields to extend the criteria of the 1st In Vivo Process. Users can indicate the relationship of the immunogen to the epitope. For example, should the immunogen be the epitope itself, the source antigen of the epitope, or have a taxonomic relationship to the epitope, such as an epitope that is reported from strain A, but the immunogen used is from strain B. These fields are mirrored in the antigen and assay fields as well.

The final specialized search is the Identifier Search, which is divided into two sections. The top section has query fields for internal IEDB identifiers for Epitope ID, Reference ID, Submission ID, and Assay ID. All curated articles and submission are given a Reference ID. In addition, data submissions are given a Submission ID that researchers can use in their publications. When the paper is published and curated, a link in the database is created between the paper and

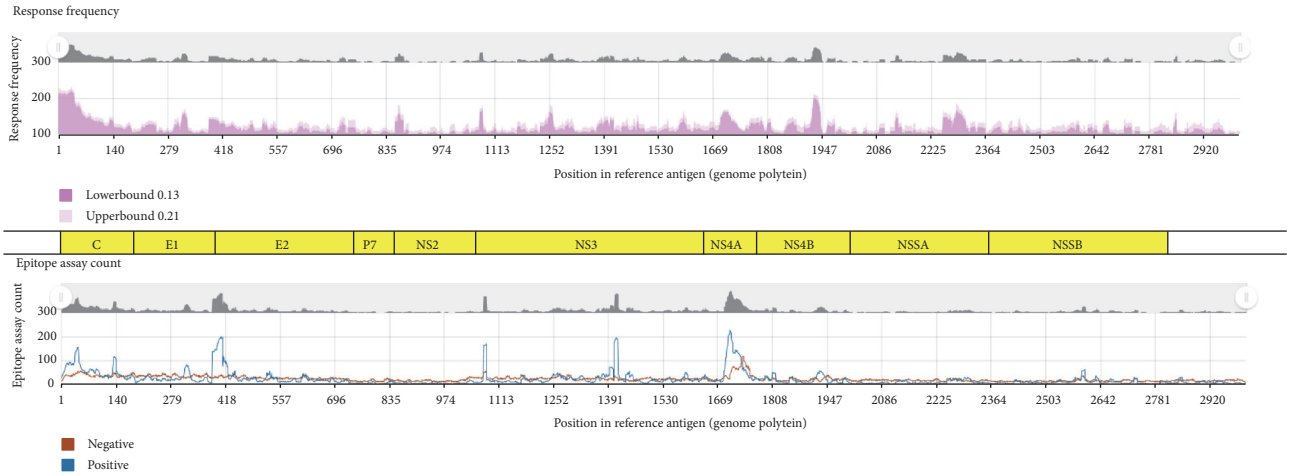


Figure 4: The Immunome Browser maps response frequencies and assay counts onto a reference antigen. The polyprotein for HCV is superimposed on the plots for reference. The data reflects IEDB HCV content as of October 2016.

corresponding submission. All assays and epitopes are also assigned unique identifiers. The bottom section of the Identifier Search contains query fields for identifiers assigned by three external databases that have links to the IEDB. These databases are PubMed, the Protein Data Bank (http://www.rcsb.org), and ChEBI.

10. The Immunome Browser Reporting Tool

The Immunome Browser (IB) is a unique analysis tool that is integrated into the query interface results page to visualize query results [17]. It displays all T cell or antibody responses along an individual protein or polyprotein/proteome. It is useful for visualizing dominant epitope regions and indicating which regions of an antigen are experimentally well-characterized. In general, it is difficult to say what an immunodominant epitope is because it is content-dependent. The IB uses host response frequency data, which is the ratio of the number of respondents to the number of subjects tested, to give an indication of the overall epitope prominence. This response frequency score (RFscore) is mapped on a residue basis onto a reference protein or proteome obtained from NCBI.

HCV provides an excellent example to demonstrate the utility of this tool because its genomic polyprotein is about 3000 amino acids in length, yet there are more than 4000 HCV epitopes reported from the literature and therefore captured in the IEDB [17]. Though seemingly counterintuitive, the large number of epitopes is explained by the fact that many variants of HCV sequences in overlapping, but not identical frames, have been tested in hundreds of different assays (T and B cell) in multiple host systems. To start, the user types "HCV" in the Organism text field in the Antigen pane on the home page. The auto-complete feature presents a list of possibilities, and the user can select the top item, Hepatitis C virus (ID:11103). Upon clicking the Search button, approximately 4500 epitopes, over 13,000 assays, and nearly 600 references are returned in the Results page. The IB is accessed from the Antigen tab. When this tab is viewed for the first time, a pop-up help window appears that points to the IB icon and states that the "Immunome Browser maps epitopes retrieved from a query onto their source protein to visualize how often different regions in a protein have been tested and how often they were positive."

Clicking on the IB icon for the HCV genomic polyprotein generates the IB page, which contains two graphs and a data table. The top graph maps the response frequency, starting at residue 1 and extending to residue 3011. Figure 4 shows the response frequency with an overlay of the individual proteins that make up the HCV genome. There is a large response for the core protein at the N-terminus and good responses for the NS4B and NS5A regions. The graph plots two lines, one in light pink and the other in dark pink. These denote the upper and lower bound, respectively, of the 95% confidence interval. The closer these two lines are to one another, the greater the confidence in the response at that residue position. Each IB plot of response frequency scores will automatically include the name if the reference proteome/genome selected for mapping of queried epitope data; however, the graphical overlay of the individual HCV proteins depicted here was superimposed within the figure as a point of reference and is not an automated feature of the IB.

The bottom graph plots the positive and negative assay counts, which is helpful in interpreting the upper graph. A region of low response frequency might reflect negative assay results or the fact that few experiments have covered that portion of the genome. The data table at the bottom of the web page (Figure 5) lists all the epitopes by position and their related data. The table contains columns for the response frequency, the number of subjects tested and responded, and the number of positive and negative assays. This table can be downloaded into Excel or other spreadsheet programs for further analysis, for example, sorting on epitopes with high RFscores.

Results Returned: 6877 | Displaying: 6877 ☐ Display Graphed Residue Positions Export Results

Epitope ID	Epitope Sequence	Mapped Position	Identity	Subjects Tested	Subjects Responded	Assays Positive	Assays Negative	Response Freq.(95% CI)
42687	MSTNPKPQKKNKRNTNRRPQDVKFPGGG	1-28	92%	1	1	1	0	1.00 (0.04:1.00)
42673	MSTIPKPQRKTKRN	1-14	92%	60	3	1	0	0.05 (0.02:0.14)
42693	MSTNPKPQRKIKRNTNRRPQDVKFPGGG	1-28	96%	1	1	1	0	1.00 (0.04:1.00)
42691	MSTNPKPQR	1-9	100%	3	0	0	1	0.00 (0.00:0.61)
42679	MSTNPKEFRKTKRN	1-14	85%	60	5	1	0	0.08 (0.04:0.18)
42682	MSTNPKPQK	1-9	88%	4	0	0	2	0.00 (0.00:0.51)
42694	MSTNPKPQRKTK	1-12	100%	11	7	1	0	0.64 (0.35:0.86)
42695	MSTNPKPQRKTKRN	1-14	100%	1	0	0	1	0.00 (0.00:0.94)
42677	MSTLPKPQRQTKRNTPRRPQNVKFPGGGQIVGGVY VLPRRGPRL	1-44	88%	1	1	1	0	1.00 (0.04:1.00)
42704	MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVY LLPRRGPRL	1-44	100%	1	1	1	0	1.00 (0.04:1.00)

Figure 5: The Immunome Browser generates a table that lists all epitopes and corresponding response information sorted by their mapped position. This table can be downloaded in a CSV-formatted file by clicking on the "Export Results" link in the upper right corner.

11. An Example Query Using the Home Page Search Interface

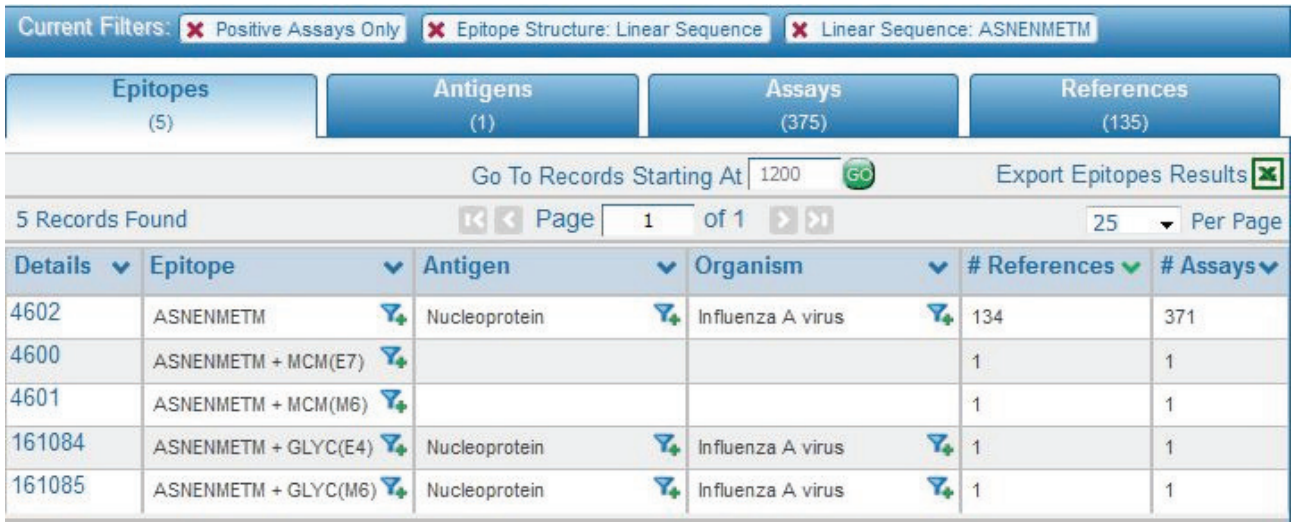
Here, we use as an example the well-characterized epitope from influenza A nucleoprotein (NP) protein, ASNENMETM. This can be done by entering or pasting the sequence in the Epitope pane sequence text field on the home page and clicking the Search button. As of January 2017, performing this query results in five epitopes, including the naturally occurring sequence from influenza A virus, and four entries representing that epitope with different posttranslational modifications. One of them, ASNENMETM + MCM(E7), indicates the main chain modification of residue seven. Another, ASNENMETM + GLY(E4), indicates a glycosylation of residue four. The sequence was identified in 375 different assays described in 135 references, as listed on the Assay and Reference tabs, respectively, as seen in Figure 6. The Assay tab itself has three tabs for T cell, B cell, and MHC Ligand assays (binding and elution). Most of the assays from this query are T cell (325), with the others being MHC ligand assays (50). Each row in the Assays tab result table represents the elements for one assay. The columns indicate the reference, epitope, host, immunization process, the assayed antigen, its relationship to the epitope, the MHC restriction (if defined and recorded), the assay type, and the qualitative value (positive or negative). By clicking on the Assay ID in the first column, users can access the relevant Assay Details page, in order to see the full curation. The Reference tab provides a summary of the articles and submissions that were curated. As mentioned above, the Ref ID is a unique internal identifier assigned by the IEDB to journal articles and data submissions. There is also a column for PubMed IDs, which is linked to the relevant entry in PubMed.

On the results page, users can take advantage of the various filter functions embedded throughout the tabs. For example, each epitope record in the Epitopes tab has a funnel icon to the right of the sequence. Clicking on this icon redisplays the query results to show data only for that epitope. In the case of the epitope ASNENMETM with Epitope ID 4602, clicking on the filter icon generates results with only one epitope, one antigen, 371 assays, and 134 references, slightly less data than the original query since the posttranslational modifications are omitted.

Users can also search for epitopes that are similar. On the home page, in addition to typing "ASNENMETM" in the Epitope pane text field, the user can use the pull-down menu to the left of the text field to select three different levels (70%, 80%, and 90%) for a BLAST match of the sequence. Setting the level to 70% and clicking the Search button for ASNENMETM yields 176 epitopes. Many of these are naturally occurring and are variations of the nucleoprotein (NP) of influenza A virus from different strains. The results table in Figure 6 also contains epitopes with no antigen or organism listed. These are termed analogs as their amino acid sequence does not have a natural source. By clicking on the filter icon for one of these analogs, Epitope ID 733, and then viewing the Assays tab of the new results, the user sees there are four assays, two T cell, and two MHC ligand. The user can find more information about each assay by clicking on the Assay ID in the leftmost column. The Assay details page contains all the information on that particular experiment, including reference (journal article or data submission), epitope, host, immunization, assay type, and antigen. In the case of Assay ID 1004030, the details page states that it is an analog in the Related Object Type row. Other details include describing the assay as a chromium release, the host as a C57BL/6 mouse, the in vitro administration as involving restimulation in vitro with effector cells from the spleen, and the antigen presenting cells as EL-4.

12. Examples That Highlight Additional Query and Reporting Capabilities

The first example provides insight into the utility of downloading query results with the Excel export option, using a query for all T cell epitopes for Dengue virus. It can be executed from the home page by typing "dengue" in the Antigen pane's Organism text field and selecting "Dengue virus (ID: 12637)" from the auto-complete list, and then by unchecking



Details	Epitope	Antigen	Organism	# References	# Assays
4602	ASNENMETM	Nucleoprotein	Influenza A virus	134	371
4600	ASNENMETM + MCM(E7)			1	1
4601	ASNENMETM + MCM(M6)			1	1
161084	ASNENMETM + GLYC(E4)	Nucleoprotein	Influenza A virus	1	1
161085	ASNENMETM + GLYC(M6)	Nucleoprotein	Influenza A virus	1	1

Figure 6: The results for a home page query for the linear sequence ASNENMETM, a well-characterized epitope of influenza A virus nucleoprotein. The results include four epitopes with posttranslational modifications.

the B Cell Assays and MHC Ligand Assays boxes in the Assay pane before clicking the Search button. As of January 2017, this query yields 1940 epitopes, 2 antigens, 3506 assays, and 85 references in the results set. Going to the Assay tab, the user can click the Excel icon to download data and further analyze it. The download of data from the Assays tab will provide a complete set of data, including details for the epitope, immunization processes, immunogens, antigens, all assay types, and reference data. The Assay tab itself contains three subtabs, one for each general assay type, and the assay count is listed at the top of the tab in parentheses.

The downloaded file is in a comma-separated value (CSV) format that can be opened in Excel or other spreadsheet programs. It contains numerous columns, each one representing a data field. The first nine columns describe the reference, followed by twelve columns describing the epitope. Other groups include host, in vivo processes, assay, antigen, MHC allele, and more. Users can sort and filter the data and utilize a variety of Excel functions to manipulate the data and perform analyses.

The second example involves using the protein tree within the Antigen finder to query for envelope proteins of several types of *Flavivirus*. Clicking on the Search button on the home page search interface leads to the results page where the antigen finder can be accessed in the Antigen pane. Clicking the blue Finder button next to Antigen Name opens this feature. By typing "flavivirus" in the Source Organism text field and selecting "Flavivirus (ID: 11051)" from the auto-complete list and clicking the Search button returns search results that appear in the bottom section of the finder. Clicking the yellow highlight icon next to "Flavivirus protein" expands the protein node of the molecule finder tree. In this case, family Flaviviridae is a high node on the tree and the first branch is Flavivirus protein representing a genus under that node. Branches within that node represent proteins from different Flavivirus species (dengue, JEV, etc.). The user can expand the Dengue virus protein nodes to reveal the envelope protein node and repeat this process to select envelope proteins from Japanese encephalitis virus, West Nile virus, Yellow fever virus, and so on, to examine responses to this protein across species. The results can also be visualized in the Immunome Browser.

13. Example Queries Using Specialized Searches

The specialized search interfaces enable more complex and intricate queries than can be performed with the home page search interface. For example, the T Cell Details Specialized search can be used to find Rift Valley fever virus T cell epitopes testing specifically on $CD8^+$ T cells and defined in humans. This search may specifically query for experiments where the authors purify and test $CD8^+$ T cells. To perform this search, the user would select T Cell Assays Details from the Specialized Searches pull-down menu and then perform the following actions. In the Epitope pane on the left, type "rift" in the Organism text field and select "Rift Valley fever virus (ID: 11588)" from the auto-complete list. In the Host pane, type "human" in the Host Organism text field and select "Homo sapiens (human) (ID: 9606, human)" from the auto-complete list. In the Assay pane, expand the Effector Cells section and select "T cell CD8+" from the multiselect pull-down menu for Effect Cell Type. Upon clicking the Search button, the results are displayed in the same format as appears for the home page search, showing 61 epitopes and 73 assays.

A user might want to further refine the query results to find out if those CD8+ T cell epitopes have been shown to work as tetramers or if tetramers have been made and shown to be effective. The query above can be refined by typing "tetramer" in the Assay text field in the Assay pane and selecting "qualitative binding/multimer/tetramer" from the auto-complete list. The Search button on the page will have changed from grey to green to indicate that it must be clicked to reinitialize the search. The refined search results show that the IEDB contains two epitopes and two positive tetramer assays.

The second specialized search example uses the B Cell Details search to find Influenza A virus epitopes recognized by neutralizing human monoclonal antibodies. Starting with the B Cell Details search page, the user would specify organism as Influenza A in the Epitope pane and host organism as human in the Host pane. For the specialized searches, the qualitative assay is not set to positive by default, so it would be necessary to select "positive" in the pull-down menu in the Assay pane for Qualitative Measurement. Next, the user would perform the following actions: in the same pane, type "neutralization" in the Assay text field and select "neutralization|biological activity" from the auto-complete list. To specify monoclonal antibodies, expand the Assayed Antibody section on the Assay pane and select "Monoclonal" from the drop-down multiselect list under Antibody Purification Status. Other choices in this menu include polyclonal and display library. Clicking the Search button reveals 66 epitopes and 575 B cell assays.

14. Example Queries for DENV Monoclonal Antibodies, Influenza Tetramer Assays, and RSV Protective Antibodies

The next query example is what monoclonal antibodies have been described for Dengue virus, which requires the B cell assay specialized search. Here, the user would perform the following actions: in the Assay pane, expand "Assay Antibody," open the menu for Antibody Purification Status, and check "monoclonal." "Display Library (monoclonal)" can also be selected if desired. In the Epitope pane, type "dengue" in the Organism text box and click on "Dengue virus (ID: 12637)," and then click the Search button to generate a list of all monoclonal antibodies for Dengue. To further explore the results, users can go to the Assay tab and download the results in an Excel-readable file. Once in Excel, users can easily sort and filter data. For example, users can filter on the Assay/Assay Group column for neutralization assays and further filter on the 1st in vivo Process column for Occurrence of Natural Infection. In addition, information is available on isotype and other assay variables. If users are interested in 3-dimensional structures, they can open the Assay finder in the Assay pane and click on "3D structure." If users are interested in a particular PDB ID, they can search for it by expanding the 3D Structure of Complex in the Assay pane and enter the PDB ID in the corresponding text field.

Other example queries for tetramers are used in the context of Influenza A virus and associated MHC alleles. To specify an assay type, the user will need to access an assay finder. Starting on the home page, the user immediately clicks the Search button to get to the Results page. In the Organism text field in the Antigen pane, the user would type "influenza" and the auto-complete function will list "Influenza A virus (ID: 11320)." Next, the user would perform the following actions: select this value and then deselect B Cell Assays and MHC Ligand Assays in the Assay pane. In that same pane, type "tetramer" in the T Cell Assays text field. From the auto-complete-generated list, select "qualitative binding|multimer/tetramer" assay and click the Search button. This will produce a list of over 400 epitopes. The user can view the alleles on the Assays tab in the MHC Restriction column, which is sortable. Users can also download the results of the Assay tab and find the allele information in the MHC/Allele Name column in one of the far right columns in the spreadsheet. There is also a column for Effector Cells/Cell Type that can be sorted and filtered.

The final infectious disease query example involves finding protective antibody epitopes for respiratory syncytial virus (RSV). This search involves targeting a specific type of assay for a pathogen of interest. Thus, from the home page, the user clicks the Search button to get to the filters on the Results page. In the Antigen pane, typing "RSV" will generate an auto-complete list with numerous related individual species choices. Thus, in order to select a high node including all RSVs, the user can instead open the finder and type "rsv" in the organism name text box. Once done, the user can click on the yellow highlight icon in one of the rows to open up the organism tree to reveal bovine RSV, human RSV, and other related organisms under the Orthopneumovirus node. From here, the user can select this higher node and click Apply. The user next specifies the assay type in the finder in the Assay pane by typing "challenge" in the name field and clicking Search to get a list of challenge assays. By clicking on the highlight icon in one of the finder search result rows, the assay tree is opened to reveal the hierarchical structure. Next, the user would perform the following actions: select the "challenge" node, click Apply, and then click Search on the Results page to reveal at least 17 epitopes in the context of an in vivo challenge assay where, in most instances, an animal model was challenged with live virus to demonstrate whether or not an epitope could elicit protective immunity.

15. Example Queries for Human Diabetes Epitopes

The first query seeks to specifically identify all human data available for type 1 diabetes (T1D), including what antigens have been mapped for T cell epitopes, and how many therapeutic epitopes have been identified. The query can be performed with the home page search by clicking Search to get to the Results page and then perform the following actions: in the Disease pane, type "diabetes" in the Specific Disease text field and select "diabetes mellitus" from the auto-complete list. In the Host pane, select the "human" radio button and in the Assay pane, uncheck B Cell Assays and MHC Ligand Assays, and click the Search button. The results of the search show over 450 epitopes. The Antigens tab displays the antigens in descending order of the number of epitopes. In this case, Gluctamate decarboxylase 2 (GAD), one of the predominant antigens for type 1 diabetes, appears at the top of the list.

Next, to find the subset of these data that is associated with treatments or decreased disease, the user can refine the initial search by adding the criteria of T cell assays related to treatment. In the T Cell Assay text field, type "treatment" and select "decreased disease in vivo assay (reduction of disease after treatment)" and click the Search button. As of

January 2017, this search revealed two epitopes found to be therapeutic in humans with T1D. The final example query builds on the previous one, asking if there are immunodominant regions defined for insulin in humans with T1D. Perform the query as above for humans, T1D and T cell responses. Then, on the antigen tab, click the Immunome Browser icon next to insulin. This will display the response frequency of all assays mapped onto a reference antigen for insulin. From this, the user can visualize the immunological hot spots.

16. Conclusions

The IEDB provides biomedical researchers interested in the development of vaccines and therapeutics a unique resource for epitopes related to a wide variety of infectious diseases, allergens, autoimmune diseases, and alloantigens. The database content of peptidic, conformational, and nonpeptidic epitopes and their associated T cell, B cell, MHC binding, and MHC ligand elution assays is kept current with the scientific literature and includes data depositions from the research community. All data in the IEDB are experimentally derived, and it contains both positive and negative data, an important consideration for experiment design and for those interested in developing machine learning algorithms for predicting epitopes.

The IEDB has certain constraints and limitations. The IEDB does not curate HIV or cancer references, as these data are not within the programmatic scope of the NIAID. However, exceptions for the inclusion of HIV and cancer epitopes are made if they are reported alongside other epitopes. Similarly, any epitope reported in the context of a cancer resulting from viral infection (e.g., human papilloma virus and human T-lymphotropic virus) are included. The database is also limited by the information provided in each paper by the authors. For example, the type of disease (disease state) is captured in accordance with the patient histories provided by the author. The IEDB primarily describes disease states using the Disease Ontology (DO), which uses formal nomenclature and synonyms. The progress of the disease (disease stage) is described using a searchable list of terms, such as acute, chronic, and post.

There are several user interfaces that researchers can use to query the database. The home page search interface features six query criteria used in most searches. Search results can be further refined with additional filters, similar to the search paradigm used on most travel websites and many shopping websites. The results page includes four separate tabs that list epitopes, antigens, assays, and references (journal articles and data submissions). A visualization of the results mapped on to a reference proteome called the Immunome Browser can be accessed from the Antigen tab. Users can also access all the fields in the data schema with the specialized search interfaces. These interfaces enable database queries that would not be possible via the home page interface. All results can be downloaded as a CSV file for further processing in a spreadsheet application. Examples are presented to illustrate the features of the home page and specialized search interfaces for a variety of applications, including infectious and autoimmune diseases.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgments

This work was supported by the National Institute of Allergy and Infectious Diseases (Contract no. HHSN272201200010C).

References

[1] K. Yusim, B. T. Korber, C. Brander, B. F. Haynes, R. Koup, J. P. Moore, B. D. Walker, and D. I. Watkins, Eds., *HIV Molecular Immunology 2009*, Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico, 2009, LA-UR 09-05941, http://www.hiv.lanl.gov/content/immunology/pdf/2009/immuno2009.pdf.

[2] R. Vita, J. A. Overton, J. A. Greenbaum et al., "The immune epitope database (IEDB) 3.0," *Nucleic Acids Research*, vol. 43, no. D1, pp. D405–D412, 2015.

[3] R. Vita, K. Vaughan, L. Zarebski et al., "Curation of complex, context-dependent immunological data," *BMC Bioinformatics*, vol. 7, no. 1, p. 341, 2006.

[4] Z. Lu and L. Hirschman, "Biocuration workflows and text mining: overview of the BioCreative 2012 workshop track II," *Database: The Journal of Biological Databases and Curation*, vol. 2012, p. bas043, 2012.

[5] R. Vita, L. Zarebski, J. A. Greenbaum et al., "The immune epitope database 2.0," *Nucleic Acids Research*, vol. 38, Database issue, p. D854, 2010.

[6] B. Smith, W. Ceusters, B. Klagges et al., "Relations in biomedical ontologies," *Genome Biology*, vol. 6, no. 5, p. R46, 2005.

[7] T. Madden, "The BLAST sequence analysis tool. 2013 Mar 15," in *The NCBI Handbook [Internet]. 2nd edition*, J. McEntyre and J. Ostell, Eds., National Center for Biotechnology Information (US), Bethesda (MD), 2013, Available from: http://www.ncbi.nlm.nih.gov/books/NBK153387/.

[8] P. de Matos, A. Dekker, M. Ennis et al., "ChEBI: a chemistry ontology and database," *Journal of Cheminformatics*, vol. 2, Supplement 1, p. P6, 2010.

[9] R. Vita, B. Peters, Z. Josephs et al., "A model for collaborative curation, the IEDB and ChEBI curation of non-peptidic epitopes," *Immunome Research*, vol. 7, no. 1, pp. 1–8, 2011.

[10] W. A. Kibbe, C. Arze, V. Felix et al., "Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data," *Nucleic Acids Research*, vol. 43, Database issue, pp. D1071–D1078, 2015.

[11] A. Bandrowski, R. Brinkman, M. Brochhausen et al., "The ontology for biomedical investigations," *PloS One*, vol. 11, no. 4, article e0154556, 2016.

[12] R. Vita, J. A. Overton, E. Seymour et al., "An ontology for major histocompatibility restriction," *Journal of Biomedical Semantics*, vol. 7, no. 1, p. 1, 2016, eCollection 2016.

[13] X. T. Wee, Y. Koh, and C. W. Yap, "Automated systems to identify relevant documents in product risk management," *BMC Medical Informatics and Decision Making*, vol. 12, no. 1, p. 13, 2012.

[14] P. Wang, A. A. Morgan, Q. Zhang, A. Sette, and B. Peters, "Automating document classification for the immune epitope database," *BMC Bioinformatics*, vol. 8, no. 1, p. 269, 2007.

[15] E. Seymour, R. Damle, A. Sette, and B. Peters, "Cost sensitive hierarchical document classification to triage PubMed abstracts for manual curation," *BMC Bioinformatics*, vol. 12, no. 1, p. 482, 2011.

[16] R. Vita, B. Peters, and A. Sette, "The curation guidelines of the immune epitope database and analysis resource," *Cytometry. Part a*, vol. 73, no. 11, pp. 1066–1070, 2008.

[17] Y. Kim, K. Vaughan, J. Greenbaum, B. Peters, M. Law, and A. Sette, "A meta-analysis of the existing knowledge of immunoreactivity against hepatitis C virus (HCV)," *PloS One*, vol. 7, no. 5, article e38028, 2012.