

REVERSE VACCINOLOGY

EDITED BY: Pedro A. Reche, Richard Moxon and Rino Rappuoli
PUBLISHED IN: *Frontiers in Immunology*





frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-385-2

DOI 10.3389/978-2-88963-385-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

REVERSE VACCINOLOGY

Topic Editors:

Pedro A Reche, Complutense University of Madrid, Spain

Richard Moxon, University of Oxford, United Kingdom

Rino Rappuoli, GlaxoSmithKline (Italy), Italy

Citation: Reche, P. A., Moxon, R., Rappuoli, R., eds. (2020). Reverse Vaccinology. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-385-2

Table of Contents

05	<i>Editorial: Reverse Vaccinology</i>
	Richard Moxon, Pedro A. Reche and Rino Rappuoli
07	<i>Trans-Allelic Model for Prediction of Peptide:MHC-II Interactions</i>
	Abdoelnaser M. Degoot, Faraimunashe Chirove and Wilfred Ndifon
18	<i>Cross-Reactive Bactericidal Antimeningococcal Antibodies can be Isolated From Convalescing Invasive Meningococcal Disease Patients Using Reverse Vaccinology 2.0</i>
	Fadil A. Bidmos, Simon Nadel, Gavin R. Screaton, J. Simon Kroll and Paul R. Langford
27	<i>A Web Resource for Designing Subunit Vaccine Against Major Pathogenic Species of Bacteria</i>
	Gandharva Nagpal, Salman Sadullah Usmani and Gajendra P. S. Raghava
38	<i>Bacterial Vaccine Antigen Discovery in the Reverse Vaccinology 2.0 Era: Progress and Challenges</i>
	Fadil A. Bidmos, Sara Siris, Camilla A. Gladstone and Paul R. Langford
45	<i>A Review on T Cell Epitopes Identified Using Prediction and Cell-Mediated Immune Models for Mycobacterium tuberculosis and Bordetella pertussis</i>
	Yuan Tian, Ricardo da Silva Antunes, John Sidney, Cecilia S. Lindestam Arlehamn, Alba Grifoni, Sandeep Kumar Dhanda, Sinu Paul, Bjoern Peters, Daniela Weiskopf and Alessandro Sette
59	<i>Trained Immunity-Based Vaccines: A New Paradigm for the Development of Broad-Spectrum Anti-infectious Formulations</i>
	Silvia Sánchez-Ramón, Laura Conejero, Mihai G. Netea, David Sancho, Óscar Palomares and José Luis Subiza
70	<i>Genome-Based Approach Delivers Vaccine Candidates Against Pseudomonas aeruginosa</i>
	Irene Bianconi, Beatriz Alcalá-Franco, Maria Scarselli, Mattia Dalsass, Scilla Buccato, Annalisa Colaprico, Sara Marchi, Vega Masignani and Alessandra Bragonzi
79	<i>The Impact of Nucleotide Sequence Analysis on Meningococcal Vaccine Development and Assessment</i>
	Martin Christopher James Maiden
91	<i>Comparison of Open-Source Reverse Vaccinology Programs for Bacterial Vaccine Antigen Discovery</i>
	Mattia Dalsass, Alessandro Brozzi, Duccio Medini and Rino Rappuoli
103	<i>The Development of a Vaccine Against Meningococcus B Using Reverse Vaccinology</i>
	Vega Masignani, Mariagrazia Pizza and E. Richard Moxon
117	<i>Comprehensive Evaluation of the Expressed CD8+ T Cell Epitope Space Using High-Throughput Epitope Mapping</i>
	Paul V. Lehmann, Maneewan Suwansaard, Ting Zhang, Diana R. Roen, Greg A. Kirchenbaum, Alexey Y. Karulin, Alexander Lehmann and Pedro A. Reche

130 Application of Modeling Approaches to Explore Vaccine Adjuvant Mode-of-Action

Paul R. Buckley, Kieran Alden, Margherita Coccia, Aurélie Chalon, Catherine Collignon, Stéphane T. Temmerman, Arnaud M. Didierlaurent, Robbert van der Most, Jon Timmis, Claus A. Andersen and Mark C. Coles



Editorial: Reverse Vaccinology

Richard Moxon^{1*†}, Pedro A. Reche^{2*†} and Rino Rappuoli^{3,4*†}

¹ Department of Paediatrics, University of Oxford, Oxford, United Kingdom, ² Department of Immunology & O2, Complutense University of Madrid, Madrid, Spain, ³ GSK, Siena, Italy, ⁴ Faculty of Medicine, Imperial College, London, United Kingdom

Keywords: infectious diseases, vaccines, reverse vaccinology, microbiology, vaccinology

Editorial on the Research Topic

Reverse Vaccinology

For many, the semantics of the term “reverse vaccinology” may be puzzling. Literally, it implies a complete change of direction or action in the study of vaccines. The non-obvious point is that this *volte face* first came about through whole genome sequencing (WGS). WGS revolutionized biology, including microbiology. Specifically, it introduced a top-down, computer data-based approach to the discovery of candidate vaccine antigens; highly sensitive, but not specific and, crucially, not hypothesis driven. This contrasted with the classical laboratory based, hypothesis driven analysis of microbes to identify components that could elicit protective immunity. *Reverse vaccinology* relies on the use of computational methods and tools to identify vaccine candidates for further experimentation, refinement of which is crucial for their optimal use as argued and detailed by Dalsass et al. These computational tools serve to anticipate antigens that are likely to induce protective responses as well as the precise antigen regions, epitopes, recognized by the immune system (1).

Reverse vaccinology was first used to predict potential antigens for a vaccine against the B strains of *Neisseria meningitidis* (meningococci) in the 1990's, reviewed by Masignani et al. It is worth emphasizing that the formulation of this complex vaccine could not have been achieved without the systematic, WGS-based approach to population biology succinctly captured in the review by Maiden, a pioneer in laying the foundations of the molecular epidemiological tools so crucial to the design of vaccines both for infectious and non-infectious diseases. More recently, Bianconi et al. have been successful in applying the classical *reverse vaccinology* approach for *Pseudomonas aeruginosa*. Starting from the 5,570 open reading frames of the genome, they selected 52 vaccine candidates by applying a number of filters to exclude the proteins predicted not to be present on the bacterial surface, to be variable in different strains, to have homology to human proteins, or to be homologous to *E. coli* proteins. Of the 52 predicted vaccine antigens, 30 were successfully expressed and several of those gave a quite remarkable protection in the mouse challenge model. However, one of the main aims of this current series on *reverse vaccinology* is to highlight how many new concepts and technologies have been recruited to facilitate vaccine design including contributions from proteomics, immunology, structural biology, systems biology, and mathematical modeling. Thus today, the change of direction and action in vaccine research, captured in the term *reverse vaccinology*, embodies much more than innovation in antigen discovery.

Bidmos et al. describe the isolation and recombinant expression of the variable regions of heavy (VH) and light (VL = κ or λ) chain genes of immunoglobulin (IgG) using a variety of molecular tools. Referred to as *reverse vaccinology 2.0*, this permits the high-throughput screening of large numbers of antibody-secreting cells and was employed to identify functional anti-*Staphylococcus aureus* monoclonal antibodies induced during bacteremia (2) and anti-MTB surface antigen antibodies cloned from patient-derived plasmablasts of reactivated memory B-cell origins (3). In a further study, Bidmos et al. describe successful efforts to utilize the *reverse vaccinology 2.0* approach to identify novel functional meningococcal antigens with the potential to expand the coverage of currently licensed meningococcal B vaccines.

OPEN ACCESS

Edited and reviewed by:

Denise Doolan,

James Cook University, Australia

*Correspondence:

Richard Moxon

richard.moxon@paediatrics.ox.ac.uk

Pedro A. Reche

parecheg@med.ucm.es

Rino Rappuoli

rino.rappuoli@gsk.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Vaccines and Molecular Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 03 November 2019

Accepted: 13 November 2019

Published: 03 December 2019

Citation:

Moxon R, Reche PA and Rappuoli R
(2019) Editorial: Reverse Vaccinology.
Front. Immunol. 10:2776.
doi: 10.3389/fimmu.2019.02776

The synergism of immune-information and systems immunobiology with WGS provides crucial tools that consider not just the challenges of the identification and molecular diversity of target antigens, but the importance of expression levels and how these variables, along with host genetic variation, impact on B-cell immune responsiveness. Immunologists do their best to identify the optimal epitopes of antigens for candidate vaccines, as exemplified by the work of Nagpal et al. These authors applied an immunoinformatic pipeline that led to identifying epitope-based vaccine candidates against 14 pathogenic bacteria and made them available through a web-resource named VacTarBac. Bacteria are complex pathogens encompassing numerous protein antigens that when targeted for epitope prediction will result in a huge number of candidates. But, this plethora of information and the challenge of what can be reasonably subjected to further rigorous investigation is a daunting challenge. Thereby, to simplify further experimental advances, the authors implemented a system to identify and prioritize virulence factors or other essential genes required for pathogenicity while also discarding epitopes cross-reactive with self-proteins. The application of stringent prioritization criteria to the selected 14 pathogenic bacteria led to the identification of just 252 unique B-cell and T-cell epitopes.

T-cell epitopes can be predicted starting from WGS. Tian et al. show how they made a full map of the T-cell epitopes starting from the 4,000 open reading frames of *Mycobacterium tuberculosis*. A metric (immunogenicity score) was devised based on predictions of their immunodominance, promiscuity, HLA restriction and conservation. In a second example, they describe how the prediction of the T-cell epitopes of *Bordetella pertussis* antigens, not just those included in currently licensed acellular vaccines, may help to design novel formulations based on Th1 and Th17 immunity to overcome the limitations of the existing vaccines which induce mostly a Th2 based immunity. Degroot et al. describe a new method to predict peptide binding to major histocompatibility complex class two (MHC-II) molecules, which is the main basis to anticipate CD4 T cell epitopes. The method is based on structural analyses of peptide-MHC II interactions and can predict peptide binding for all three human MHC-II loci (HLA-DR, HLA-DP, and HLA-DQ). The authors report that the performance of the method is in general

comparable to neural network methods and is superior in predicting peptide binding to HLA-DP molecules. The main advantage of this approach reported over other machine learning models is that of being rooted on actual physicochemical peptide-MHC-II binding interactions. A main handicap is however that the authors have not made available the method for rigorous independent comparisons.

Sánchez-Ramón et al. makes a well-argued case for trained-immunity based vaccines (TibV). These are vaccines that induce an innate, non-specific immunity for long periods of time. A typical example of a TibV Vaccine is BCG which induces two types of immunity, one based on adaptive immunity specific for *Mycobacterium tuberculosis*, and the other based on innate immunity which is non-specific but so effective that it is also recommended to cure prostate cancer. This immunity induces activation of dendritic cells, activation of non-specific effector responses of innate immune cells such as monocytes and macrophages and is maintained overtime by epigenetic changes. Vaccine adjuvants and non-toxic derivatives of toxins (4) inducing non-specific protection against bacteria or viruses can be considered a proxy of TibV.

Buckley et al. describe modeling approaches that provide exciting insights into AS01, one of the most successful adjuvants licensed for human use, and how such an adjuvant may work. According to Sánchez-Ramón et al., in addition to adaptive immune responses, it is also likely to induce trained immunity. AS01 has been licensed for the RTS-S vaccine against malaria and the Shingrix vaccine against Shingles and is part of the first successful clinical trial showing protection from disease in people infected by *Mycobacterium tuberculosis*.

A quarter of a century after WGS revolutionized biology, this series is a timely and exciting opportunity to reflect on what has been achieved by *reverse vaccinology* and how best to galvanize future efforts to improve global public health through rigorous and imaginative exploitation of the explosion in technologies that can be used to develop a broad range of novel vaccines.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

1. Sanchez-Trincado JL, Gomez-Perosanz M, Reche PA. Fundamentals and methods for T- and B-cell epitope prediction. *J Immunol Res.* (2017) 2017:2680160. doi: 10.1155/2017/2680160
2. Lu DR, Tan YC, Kongpachith S, Cai X, Stein EA, Lindstrom TM, et al. Identifying functional anti-*Staphylococcus aureus* antibodies by sequencing antibody repertoires of patient plasmablasts. *Clin Immunol.* (2014) 152:77–89. doi: 10.1016/j.clim.2014.02.010
3. Zimmermann N, Thormann V, Hu B, Köhler AB, Imai-Matsushima A, Loch C, et al. Human isotype-dependent inhibitory antibody responses against *Mycobacterium tuberculosis*. *EMBO Mol Med.* (2016) 8:1325–39. doi: 10.15252/emmm.2016.06330
4. Williams AE, Edwards L, Humphreys IR, Snelgrove R, Rae A, Rappuoli R, et al. Innate imprinting by the modified heat-labile toxin of *Escherichia coli* (LTk63) provides generic protection against lung infectious

disease. *J Immunol.* (2004) 173:7435–43. doi: 10.4049/jimmunol.173.12.7435

Conflict of Interest: RR was employed by GSK and RM holds a consultancy agreement as a scientific adviser to GSK.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Moxon, Reche and Rappuoli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Trans-Allelic Model for Prediction of Peptide:MHC-II Interactions

Abdoelnaser M. Degoot^{1,2,3*}, Faraimunashe Chirove² and Wilfred Ndifon^{1*}

¹African Institute of Mathematical Sciences (AIMS), Muizenberg, South Africa, ²School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa, ³DST-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), Gauteng, South Africa

OPEN ACCESS

Edited by:

Pedro A. Reche,
Complutense University of Madrid,
Spain

Reviewed by:

Anne Searls De Groot,
EpiVax, United States
Morten Nielsen,
Technical University of Denmark,
Denmark

*Correspondence:

Abdoelnaser M. Degoot
degoot@aims.ac.za;
Wilfred Ndifon
wndifon@aims.ac.za

Specialty section:

This article was submitted to
Vaccines and Molecular Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 06 February 2018

Accepted: 06 June 2018

Published: 20 June 2018

Citation:

Degoot AM, Chirove F and Ndifon W
(2018) Trans-Allelic Model for
Prediction of Peptide:MHC-II
Interactions.
Front. Immunol. 9:1410.
doi: 10.3389/fimmu.2018.01410

Major histocompatibility complex class two (MHC-II) molecules are trans-membrane proteins and key components of the cellular immune system. Upon recognition of foreign peptides expressed on the MHC-II binding groove, CD4⁺ T cells mount an immune response against invading pathogens. Therefore, mechanistic identification and knowledge of physicochemical features that govern interactions between peptides and MHC-II molecules is useful for the design of effective epitope-based vaccines, as well as for understanding of immune responses. In this article, we present a comprehensive trans-allelic prediction model, a generalized version of our previous biophysical model, that can predict peptide interactions for all three human MHC-II loci (HLA-DR, HLA-DP, and HLA-DQ), using both peptide sequence data and structural information of MHC-II molecules. The advantage of this approach over other machine learning models is that it offers a simple and plausible physical explanation for peptide-MHC-II interactions. We train the model using a benchmark experimental dataset and measure its predictive performance using novel data. Despite its relative simplicity, we find that the model has comparable performance to the state-of-the-art method, the NetMHCIIpan method. Focusing on the physical basis of peptide-MHC binding, we find support for previous theoretical predictions about the contributions of certain binding pockets to the binding energy. In addition, we find that binding pocket P5 of HLA-DP, which was not previously considered as a primary anchor, does make strong contribution to the binding energy. Together, the results indicate that our model can serve as a useful complement to alternative approaches to predicting peptide-MHC interactions.

Keywords: major histocompatibility complex (MHC), modeling peptide-MHC-II interactions, antigen presentation, machine learning, inverse statistical mechanics

1. INTRODUCTION

Major histocompatibility complex class two (MHC-II) molecules are surface proteins that exist on the membrane of antigen presenting cells (APCs) such as macrophages, dendritic cells, and B cells. They bind short peptide fragments derived from exogenous proteins and present them to CD4⁺ helper-T cells. Upon the recognition of foreign peptides presented by MHC-II molecules, the helper-T cells (precisely speaking, CD4⁺ effector T cells) will initiate proper adaptive immune responses, including enabling sufficient maturation of B cells and cytotoxic CD8⁺ T cells (1). Therefore, the binding of peptide to MHC-II molecules is considered to be a fundamental and pre-requisite step in the initiation of adaptive immunity (2, 3). As such, mechanistic identification of the basic determinants of peptide-MHC-II interactions presents potential for understanding the immune system's mechanisms and improving the process of designing peptide- and protein-based vaccines.

MHC genes for humans, referred to as human leukocyte antigen (HLA), are among the most polymorphic genetic elements found within a long continuous stretch of DNA on chromosome 6 (4). Such high polymorphism reflects the immense contribution of MHC molecules to the adaptive immune system and underpins their capacity to recognize a wide range of pathogens. Nonetheless, some viruses, such as hepatitis C, avian/swine influenza, and human immunodeficiency virus (HIV), undergo extensive mutations that allow them to partially escape recognition by the MHC molecules (5). MHC genes can be divided into HLA class I, II, and III. Loci corresponding to HLA class I are A, B, and C; HLA class II loci are DP, DQ, and DR; HLA class III genes encode for several other immune-related proteins and provide support for the former two classes (1, 4).

MHC-II molecules account for the likelihood of success of organ transplantation, and there are well-established associations between many disorders and particular classes of MHC-II molecules. These include the contribution of HLA-DQ genes to insulin-dependent diabetes (6); HLA-DR genes to multiple sclerosis; and narcolepsy (7) along with other autoimmune diseases resulting from degeneracy and misregulation in the process of peptide presentation (8). Moreover, genetic and epidemiological data have implicated MHC-II molecules in susceptibility to many infectious diseases such as HIV/AIDS, malaria (9), and cancer (10).

Experimental assays for prediction of peptide–MHC-II interactions are often faced with important obstacles, including substantial resources needed for laboratory work, high time, and labor demands. This is the case in particular, for experimental work aimed at finding out which promiscuous epitopes bind to specific MHC molecules, a necessary step in the design of peptide-based vaccines which protect against a broad range of pathogen variants. Computational methods, which are more efficient and less costly than biological assays, have been employed to complement these assays. Due to advances in sequencing technologies, immunological data have grown at an unprecedented pace and continue to accrue. This has been exploited in systematic computational analyses of genomes of multiple pathogens to determine which subunits might induce a potent immune response. The results have been the design and development of new vaccine candidates against HIV, influenza, and other hyper-variable viruses (11). Use of computational methods has significantly reduced experimental effort and costs by up to 85% (12).

Many immunoinformatics methods for prediction of peptide–MHC interactions, for both class I and II, have been developed based on machine learning approaches such as simple pattern motif (13), support vector machine (SVM) (14), hidden Markov model (HMM) (15), neural network (NN) models (16–18), quantitative structure–activity relationship (QSAR) analysis (19), structure-based methods, and biophysical methods (2, 20, 21; Degoot et al., unpublished). These methods can be divided into two categories, namely, intra-allele (allele-specific) and trans-allele (pan-specific) methods. Intra-allelic methods are trained for a specific MHC molecule on a limited set of experimental peptide-binding data and applied for prediction of peptides binding to that molecule. Because of the extreme polymorphism of MHC molecules, the existence of thousands of

allele variants, combined with the lack of sufficient experimental binding data, it is impossible to build a prediction model for each allele. Thus, trans-allele and general purpose (22) methods such as *MULTIRTA* (2), *NetMHCIIpan* (18), and *TEPITOPEpan* (23) have been developed using richer peptide-binding data expanding over many alleles or across species (18). Similar methods for MHC-I are also available such as *NetMHCpan* (24, 25) and *KISS* (26).

The trans-allelic models are often designed to extrapolate either structural similarities or shared physicochemical binding determinants among HLA genes, to predict affinities for alleles that are not part of the training dataset. These models generally have better predictive performance for new alleles and a wide range of potential applications compared with the intra-allelic models.

Most of the existing trans-allelic models for MHC-II are extended versions of their earlier intra-allelic counterparts: *TEPITOPEpan* (23) was extended from *TEPITOPE* (27); *MULTIRTA* (2) evolved from *RTA* (20); and the series of *NetMHCIIpans* (1.0, 2.0, 3.0, and 3.1) (17, 18, 28, 29) were generalized from the NN align (30) method. In the same vein, in this article, we present a trans-allele method, an extension of our previous method (Degoot et al., unpublished), for prediction of peptide–HLA class II interactions based on biophysical ideas.

The remarkable strength of the method presented here over other existing advanced data-driven approaches is its physical basis. We formulate the process of binding affinity between peptide and MHC-II molecule as an inverse problem of statistical physics. From the observable macroscopic (bound and unbound) states of experimental data, we compute the microscopic parameters (Hamiltonians for amino acid residues involved in the interaction) that govern the system. In fact, many problems in computational biology can be solved in such a way (31, 32), taking advantage of the availability of vast amount of genomic data and high resolution structural information. Solutions obtained using this approach are more plausible and physically interpretable than those obtained using mere sequence-based methods (2; Degoot et al., unpublished). In addition, because sparsity is a hallmark feature of biological processes, we adjust the model's parameters via incorporating an L_1 regularization term into the model. The L_1 constraint, commonly named *Lasso*, encourages sparsity and improves the predictive performance of the model on novel data.

The rest of this article is organized as follows: in Section 2.1, we describe the idea of MHC-II polymorphic residue groups, which is employed to capture structure similarity among MHC-II alleles. In Section 2.2, we define our methodology and formulate the learning function. After that we briefly describe the benchmark dataset used to test the predictive performance of the model in Section 2.3 and present the results in Section 3. Finally, in Section 3.3, we summarize and discuss our results and compare our method with the state-of-the-art method.

2. MATERIALS AND METHODS

2.1. MHC-II Polymorphic Residue Groups

Crystal structures revealed that an MHC molecule is a combination of two domains, an α helix and a β sheet, linked together to form a Y-shaped groove which is used to locate peptides, and both

domains equally contribute to the binding affinity. For HLA-I molecules, the β domain is largely conserved, and variation occurs mostly in the α domain. On the other hand, polymorphism occurs in both domains of HLA-II molecules; except for HLA-DR alleles, where the variation takes place in the β domain. In addition, the peptide-binding groove of the HLA-II is open at both ends, which allows binding peptides of variable lengths (33), ranging from 9 to 30 amino acid residues, or even an entire protein (29, 34). This is in contrast to the peptide-binding groove of the HLA-I alleles, which accommodate only short peptides of lengths ranging from 8 to 11 amino acids. This flexible constraint on peptide lengths together with its immense polymorphism, contribute to a lower predictive performance of computational methods for peptide-MHC-II interactions compared with MHC-I methods (2, 22).

The notion of MHC polymorphic residue groups, introduced by Bordner and Mittelmann (2), is based on a simple observation of an intrinsic (independent of peptide) feature of the MHC-II binding groove. Although a peptide could bind to an MHC-II molecule in various registers, due to the open-ended nature of the MHC-II binding groove, the strength of the binding affinity is primarily determined by 9 residues occupying the binding groove pockets. Interestingly, most of polymorphism in MHC-II genes occurs at these binding pockets (see the discussion in Section 3.3).

From the limited available crystallographic structural data of peptide-MHC-II complexes for a few MHC-II molecules from the Protein Data Bank (PDB) (35) (summarized in Table S1 in Supplementary Material), sets of important positions for the polymorphic residues in the binding groove that contact one or more peptide-binding cores and are within a distance of not more than 4 Å (2, 18, 36) in one or more of the MHC-II complex structures can be extracted. Then, by extrapolating the similarities among MHC molecules, their corresponding residues in different genes are determined using multiple sequence analysis (MSA) (37). Exploiting the fact that HLA-DR alleles are polymorphic only in the β domain and have the same α domain, the polymorphic residue groups for HLA-DR are extracted from its β domain sequences. Similarly, assuming sufficiency of the β domains for predicting MHC-peptide binding preferences (2) and for the sake of simplicity of the model, residue groups for HLA-DP and HLA-DQ were also extracted from the β domain.

Next, the set of polymorphic residues that always co-occur at the specified positions are clustered into the same group. The rationale of clustering polymorphic residue groups, rather than individual residues, is to avoid over-parametrization of the model. Table S2 in Supplementary Material shows such polymorphic residue groups for HLA-DRB, HLA-DP, and HLA-DQ alleles, assembled by the procedures described earlier.

2.2. Trans-Allele Model

In our previous intra-allele model (Degoot et al., unpublished) the probability of peptide $P^{(k)}$ to bind an MHC molecule $M^{(T(k))}$ was computed as follows:

$$\pi \left(P^{(k)}, M^{(T(k))} \right) = \frac{1}{1 + e^{\delta E^{(k)}}}, \quad (1)$$

where $\delta E^{(k)}$ is the change in binding energy in terms of the sum of the differences of first- and second-order Hamiltonians between

the bound and unbound states. Specifically, $\delta E^{(k)}$ is given by the following equation:

$$\delta E^{(k)} = \underbrace{\sum_{i=1}^{|P^{(k)}|} \delta H^{(1)}(a_i) + \sum_{i=1}^9 \delta H^{(1)}(b_i) + \delta S}_{\text{first-order Hamiltonians per residue-residue interactions}} + \underbrace{\sum_{i=1}^{|P^{(k)}|} \sum_{j=1}^9 \sum_{r=1}^R \delta H^{(2)}(a_{ir}^{(k)}, b_j)}_{\text{second-order Hamiltonians}}, \quad (2)$$

in which $|P^{(k)}|$ is the length of peptide k , R is the number of all possible configurations (registers) in which the peptide binds to the particular MHC molecule, and δS is the difference in entropy between the bound and unbound states.

For the trans-allele model, two changes were introduced into the second term of equation (2). First, instead of residue-residue interaction, $\delta H^{(2)}(a_{ir}^{(k)}, b_j)$, with $a_{ir}^{(k)}$ on the peptide sequence and b_j on the MHC binding pocket, we rather focus on residue-polymorphic group interaction, $\delta H^{(2)}(a_{ir}^{(k)}, g_{jn})$, where g_{jn} is residue group number n of position j as defined in Section 2.1. Next, we introduce a binary operator $T(k, j, n)$ that equals 1 if the MHC molecule type, $M^{(T(k))}$, corresponding to peptide $P^{(k)}$ contains polymorphic residue group n at the set of pre-determined positions of pocket j , and equals 0 otherwise. Hence, $\delta E^{(k)}$ is given by the following equation:

$$\delta E^{(k)} = \underbrace{\sum_{i=1}^{|P^{(k)}|} + \delta H^{(1)}(a_i) + \sum_{i=1}^9 \delta H^{(1)}(b_i) + \delta s}_{\text{first-order Hamiltonians per residue-group interactions}} + \underbrace{\sum_{i=1}^{|P^{(k)}|} \sum_{j=1}^9 \sum_{r=1}^R \sum_{n=1}^{G(j)} \delta H^{(2)}(a_{ir}^{(k)}, g_{jn}) T(k, j, n)}_{\text{second-order Hamiltonians}}, \quad (3)$$

where $G(j)$ is the number of polymorphic residue groups for binding pocket j . Column two of Table S2 in Supplementary Material shows $G(j)$, $j = 1, 2, \dots, 9$, for HLA-DR, HLA-DP, and HLA-DQ alleles.

Let Δ denote the model's parameters. Using equations (1) and (3), we formulate, through the maximum likelihood approach, the following cost function:

$$\mathcal{L}(P, M|\Delta) = \operatorname{argmin}_{\{\Delta\}} \left(\sum_{k=1}^K G^k(\Delta^k) + \lambda \mathcal{P}(\Delta) \right), \quad (4)$$

where $G^k(\Delta)$ is the empirical loss function given by the following equation:

$$G^k(\Delta) = y^k \log(\pi^k(\Delta)) + (1 - y^k) \log(1 - \pi^k(\Delta)), \quad (5)$$

and $y^k \in \{0, 1\}$ is the experimental value; $y = 1$ for binding peptides and $y = 0$ for non-binding ones. $\lambda \mathcal{P}(\Delta)$ is a regularization term with the following form:

$$\lambda \mathcal{P}(\Delta) = \lambda \|\Delta\|_1 = \lambda \sum_{i=1}^d |\Delta_i|, \quad (6)$$

where $\lambda > 0$ is a hyper-parameter and d is the dimension of parameter vector Δ , which varies depending on the type of MHC-II molecule. The L_1 constraint penalty term $\mathcal{P}(\Delta)$, also known as Lasso (38), has an important role in the model. As the model is defined on a large number of parameters ($d = 2,321$, 561, and 401 for HLA-DR, HLA-DP, and DQ molecules, respectively) a few parameters are expected to contribute to the binding affinity while the rest are expected to be noisy. Lasso has the capability to filter out the noisy parameters by inducing sparsity in the model, as it shrinks most of the parameter values to 0, and avoids data over-fitting. The hyper-parameter λ controls the degree of sparsity of the model; the larger the value of λ the more sparse the model. Equation (4) is a non-linear and non-smooth function; due to the L_1 constraint. But it is a convex function and we solved it, after quadratic approximation, by means of an iterative, cyclic coordinate descent approach using a soft-thresholding operator. This learning function takes the form of a generalized linear model and the algorithm we used to solve it is both fast and efficient. Details of this optimization method are found in Friedman et al. (39) and are summarized in the supplementary material.

2.3. Binding Affinity Dataset

The model has been developed by using both quantitative peptide-binding data and MHC-II molecule sequences. We obtained a total of 51,023 peptide-binding data for 24 HLA-DR, 5 HLA-DP, and 6 HLA-DQ from the IEDB database (40). This is a well-curated dataset and was used to develop NetMHCIIpan (18), the state-of-the-art method. The binding affinities data were given in the form of log-transformed measurements of the IC_{50} (half maximum inhibition concentration) according to the formula $1 - \log(IC_{50})/\log(50,000)$ (16). We dichotomized these data using a moderate threshold of IC_{50} 500 nM ($\equiv 0.426$ of log-transformed data). Peptides with IC_{50} less than or equal 500 nM (≥ 0.426 of log-transformed value) were considered as binders, and non-binders otherwise. This moderate threshold, which has been used in other previous methods including the state-of-the-art method (20, 29, 30, 41), allows us to make direct comparisons.

Amino acid sequences for the MHC-II alleles used in this study were obtained from the EMBL-EBI online-database (42). **Table 1** gives a summary of the peptide-binding dataset used to develop the method.

3. RESULTS

This section presents prediction results of the model obtained from the dataset of three MHC-II allotypes as described in Section 2.3. We applied a fivefold cross validation analysis to the model and compared it against its intra-allelic version (Table S3 in Supplementary Material). We also examine its predictive performance on data which were previously unseen by the model.

3.1. Performance of the Trans-Allele Model

We tested the predictive performance of the model by using fivefold cross validation. The partitioning of the data used in fivefold cross validation was previously done by Andreatta et al. (29), by clustering together peptides in a way that minimizes over-estimation of predictive performance, using the technique described by Nielsen et al. (30). **Figure 1** shows results of the test done using alleles belonging to the three MHC-II loci considered in this study. The performance was measured in terms of area under the curve (AUC) (43) values, which range between 0 and 1. The higher the AUC value the better the predictive performance of model. Values below 0.5 reflect a worse performance than a random test. The model has an excellent performance for HLA-DP molecules (average AUC value = 0.930), and a good predictive power for both HLA-DQ and HLA-DR molecules (average AUC values = 0.830 and 0.802, respectively). The surprisingly excellent performance for HLA-DP could be the result of both a higher structural similarity (see Section 3.3) and a higher number of peptides per allele for HLA-DP. Indeed, for all HLA-DP alleles, the number of available peptides exceeds the empirically required number of peptide-binding measurements (≈ 200 peptides (22)), but this is not the case for all HLA-DR alleles. HLA-DQ alleles have sufficient number of peptide measurements but these have a lower structural similarity compared with the corresponding peptides for HLA-DP alleles (see Section 3.3).

3.2. Comparing the Intra-Allele vs Trans-Allele Methods

Table S3 in Supplementary Material shows AUC values obtained with the intra-allele and trans-allele versions of the model. For the intra-alleles version, the model was evaluated on peptide-binding data corresponding to an individual allele only. On average, the performance of the trans-allele model is comparable to that of the intra-allele model for HLA-DP (0.930 vs 0.928), it is worse for HLA-DQ (0.830 vs 0.857) and it is better for HLA-DR (0.780 vs 0.771) (**Figure 2**).

These results demonstrate two important observations. First, there is a common binding preference among MHC-II loci, which is the basis of all trans-allelic models, and that has been successfully captured by the definition of MHC-II polymorphic groups for HLA-DP loci, and to a lesser extent for HLA-DQ and HLA-DR. Second, the trans-allelic model is able to extrapolate similarities among the MHC-II allotypes and achieve good predictive performance. As a result, the overall performance of the trans-allelic model is comparable to that of intra-allele model, even though the former model is applied on a much diverse set of MHC-II sequences.

A decreased performance of the trans-allelic model when compared with the intra-allelic method for HLA-DQ molecules is consistent with results reported in NetMHCIIpan (18). Here we suggest that this is probably because of the limited structural information available for HLA-DQ alleles. In fact, because of this limited structural information there are only 17 polymorphic residue groups for all the 9 binding pockets defined for HLA-DQ alleles. By contrast, there are 25 and 115 polymorphic residue groups defined for HLA-DP and HLA-DR molecules, respectively.

TABLE 1 | Overview of the MHC-II peptide-binding data utilized in this study.

Allele name	HLA-index	# of Peptides	# of Binders	% of Binders
HLA-DR molecules				
DRB1*01:01	HLA00664	7,685	4,382	57.02
DRB1*03:01	HLA00671	2,505	649	25.91
DRB1*03:02	HLA00673	148	44	29.73
DRB1*04:01	HLA00685	3,116	1,039	33.31
DRB1*04:04	HLA00689	577	336	58.23
DRB1*04:05	HLA00690	1,582	627	39.63
DRB1*07:01	HLA00719	1,745	849	48.65
DRB1*08:02	HLA00724	1,520	431	28.36
DRB1*08:06	HLA00732	118	91	77.12
DRB1*08:13	HLA00739	1,370	455	33.21
DRB1*08:19	HLA00745	116	54	46.55
DRB1*09:01	HLA00749	1,520	621	40.86
DRB1*11:01	HLA00751	1,794	778	43.37
DRB1*12:01	HLA00789	117	81	69.23
DRB1*12:02	HLA00790	117	79	67.52
DRB1*13:02	HLA00798	1,580	493	31.20
DRB1*14:02	HLA00834	118	78	66.20
DRB1*14:04	HLA00836	30	16	53.33
DRB1*14:12	HLA00844	116	63	54.31
DRB1*15:01	HLA00865	1,769	709	40.08
DRB3*01:01	HLA00887	1,501	281	18.72
DRB3*03:01	HLA00902	160	70	43.75
DRB4*01:01	HLA00905	1,521	485	31.89
DRB5*01:01	HLA00915	3,106	1,280	41.21
HLA-DP molecules				
DPA1*01:03–DPB1*02:01	HLA00517	1,404	538	38.32
DPA1*01:03–DPB1*04:01	HLA00521	1,337	471	35.23
DPA1*02:01–DPB1*01:01	HLA00514	1,399	597	42.67
DPA1*02:01–DPB1*05:01	HLA00523	1,410	443	31.42
DPA1*03:01–DPB1*04:02	HLA00522	1,407	523	37.17
HLA-DQ molecules				
DQA1*01:01–DQB1*05:01	HLA00638	1,739	522	30.02
DQA1*01:02–DQB1*06:02	HLA00646	1,629	813	49.91
DQA1*03:01–DQB1*03:02	HLA00627	1,719	386	22.46
DQA1*04:01–DQB1*04:02	HLA00637	1,701	559	32.86
DQA1*05:01–DQB1*02:01	HLA00622	1,658	549	33.11
DQA1*05:01–DQB1*03:01	HLA00625	1,689	863	51.10
Total		51,023	20,255	39.70

The first column gives the names of the 34 genes used to develop the method, distributed as 24, 5, and 6 for HLA-DR, HLA-DP, and HLA-DQ genes, respectively. The second column represents the index for each allele in the **EMBL-EBI** database (42). The third and fourth columns give the total number of peptide and the number of binder peptides, respectively, per allele. The last column shows the percentage of binder peptides. Binder peptides were identified using an IC_{50} binding cutoff of 500 nM, as in previous studies (2, 17, 18, 30). The last row presents the overall statistics for the last three columns.

Another reason for the reduction of the trans-allelic model's performance for HLA-DQ alleles is that there is a large sequence diversity of MHC-II molecules belonging to this locus. We will examine the empirical support for this assertion in Section 3.3.

3.3. Prediction on a Novel Dataset

We examined the predictive power of the model on a blind dataset- i.e., a dataset which was not used in the training phase. More precisely, to make peptide-binding predictions for a particular allele, we train the model on an entirely different allele. The allele used for training was chosen based on its similarity to the focal allele as quantified using three different metrics: nearest-neighbor, Hamming distance, and Leave-One-Out (LOO) approach.

In the nearest-neighbor approach the distance between two MHC molecules is defined (17) as follows:

$$d(A, B) = 1 - \frac{S(A, B)}{\sqrt{S(A, A) S(B, B)}} \quad (7)$$

in which $S(A, B)$ is the score of the BLOSUM50 (44) metric between amino acid sequences of A and B . The BLOSUM50 metric measures genetic distance between two sequences by quantifying the likelihood that one amino acid will be substituted by another amino acid on evolutionary time scales. Hamming distance simply counts the different occurrences of corresponding amino acid residues between two sequences. In both nearest-neighbor and Hamming metrics, we train the model on peptide data belonging to the corresponding nearest allele to parameterize the model, and then we assess its accuracy in terms of AUC values calculated

AUC ROC curves for the three loci

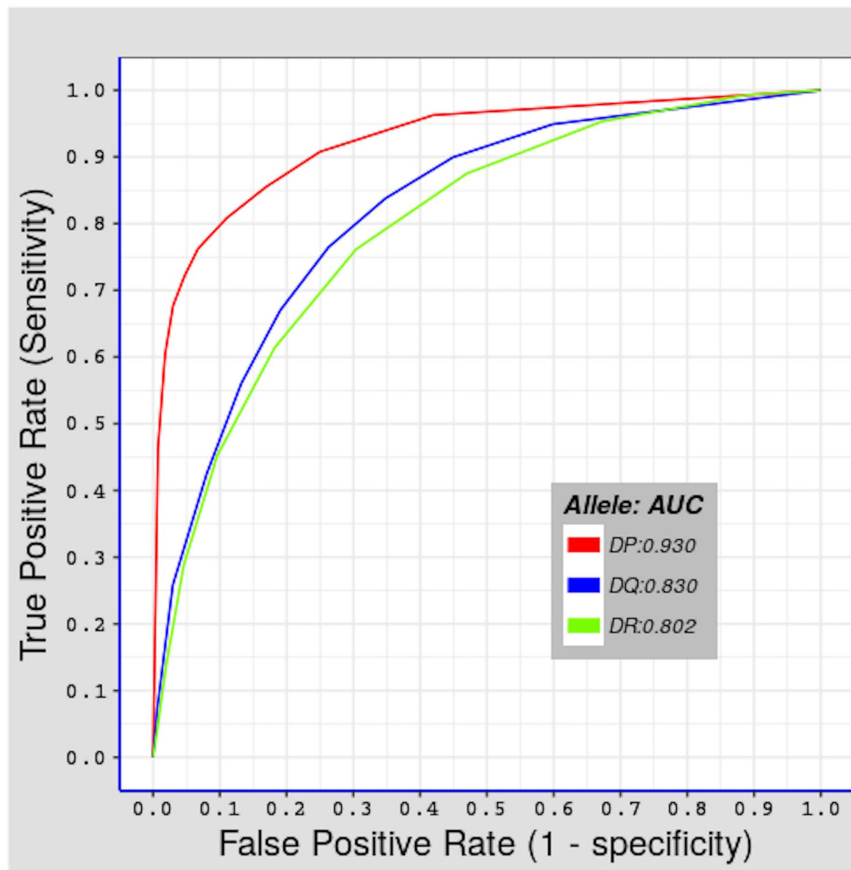


FIGURE 1 | Shows fivefold cross validation results of the model using the benchmark dataset described in Section 2.3. Three ROC curves representing the three MHC-II loci covered in this study. The red curve for HLA-DP with AUC value = 0.930, the blue curve for HLA-DQ with AUC value = 0.830, and the green curve for HLA-DR with AUC = 0.802.

based on peptide data belonging to the focal allele using those parameters.

However, unlike the TEPITOPE and the series of NetMHCIIpan methods which defined nearest neighbor at pocket level, we derive both the nearest-neighbor metric and the Hamming distance at residue level. Our choice is based on the fact that accounting for the entire MHC-II sequence provides a broader allele coverage (2) and hence extend the model's applicability. Computing sequence similarity at residue level is an intuitive and natural approach to perform comparative analysis of sequences rather than other artificial ways that may be more computationally efficient. We found that 71% (for HLA-DR), 60% (HLA-DP), and 67% (HLA-DQ) of alleles used for training were consistent between the residue-level and pocket-level approaches. These statistics indicate that, as mentioned before, most of MHC-II polymorphisms occur at the binding pockets.

The LOO approach involved partitioning data into two parts; the peptide-binding data not belonging to the allele under consideration are used to learn the model's parameters and the remaining data, the peptide-binding data belonging to the focal allele, are used as test data. **Figure 3** shows a comparison of results from these three approaches (details are in Table S4 in Supplementary Material). The results show that, regardless of the metric we

used, the trans-allele method has a high predictive power for HLA-DP allele and a moderate predictive power for the other alleles.

The much higher predictive power for HLA-DP compared with the other alleles is likely due to the comparatively lower sequence diversity of HLA-DP alleles. To make this assertion more precise we carried out a regression analysis by defining the AUC values from LOO approach as functions of both NN and Hamming metric distances. **Figure 4** gives results of our analysis. As seen in **Figure 4**, all HLA-DQ alleles fall below the least squares lines for both metrics (blue points). We also found that model performance for HLA-DP allele (red points) increases as the distance between alleles decreases. The authors of NetMHCIIpan also arrived at the same conclusion (18), but only for the NN metric.

3.4. Analysis of the Model's Parameters

To determine the key factors that contribute to the binding affinities for the three MHC-II alleles considered in this study, we calculated the Hamiltonians corresponding to each amino acid residue and the 9 binding pockets of the MHC-II binding groove. These values were then averaged over only the polymorphic residue groups defined for each pocket containing the particular amino acid.

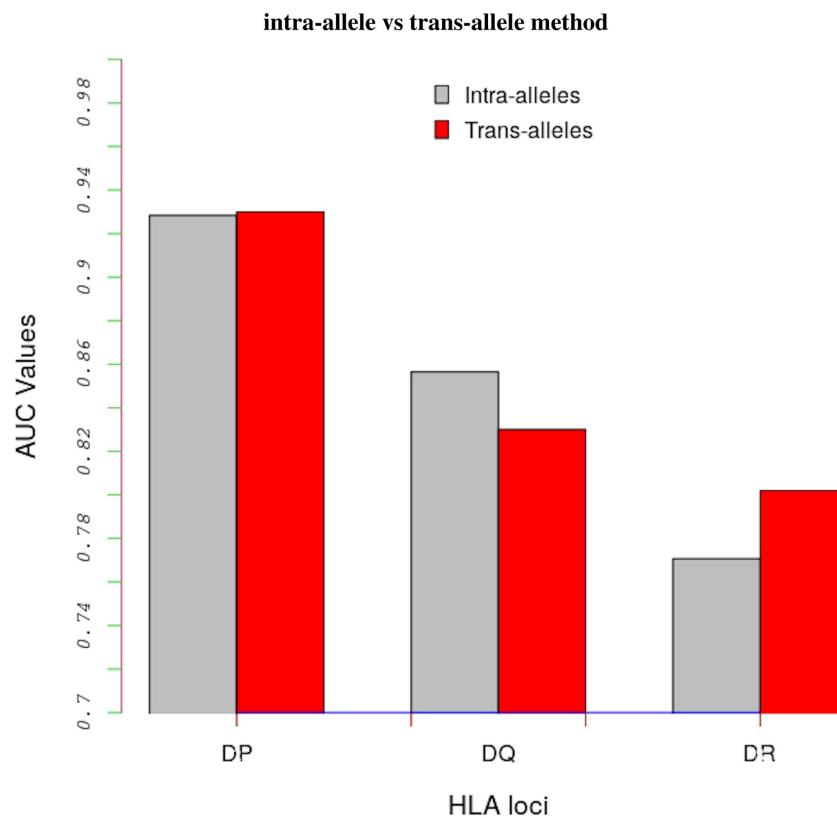


FIGURE 2 | Comparing results between the intra-alleles (gray bars) and the trans-alleles (red bars) methods in terms of AUC values. These bars show that there is a significant increase in performance of the trans-allele method for HLA-DR molecules and decrease for HLA-DQ molecules compared with the intra-allele method. The difference in the HLA-DP loci is limited.

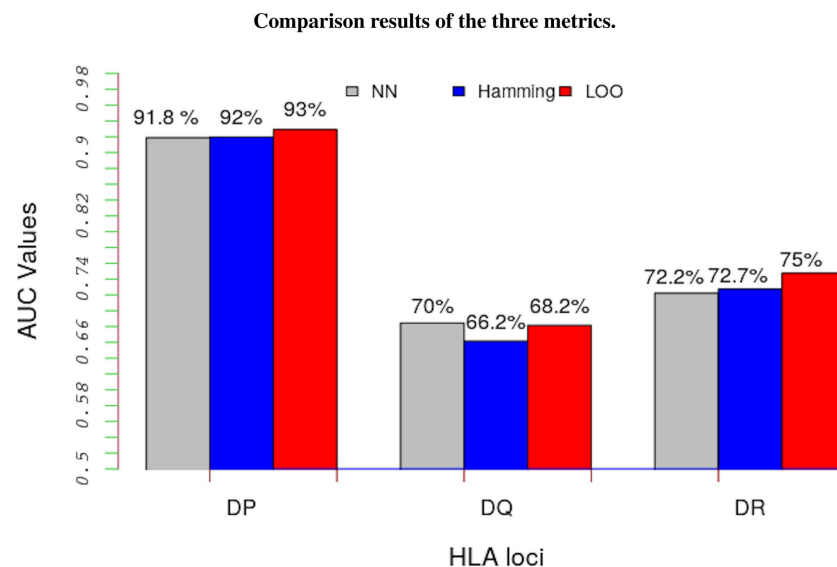
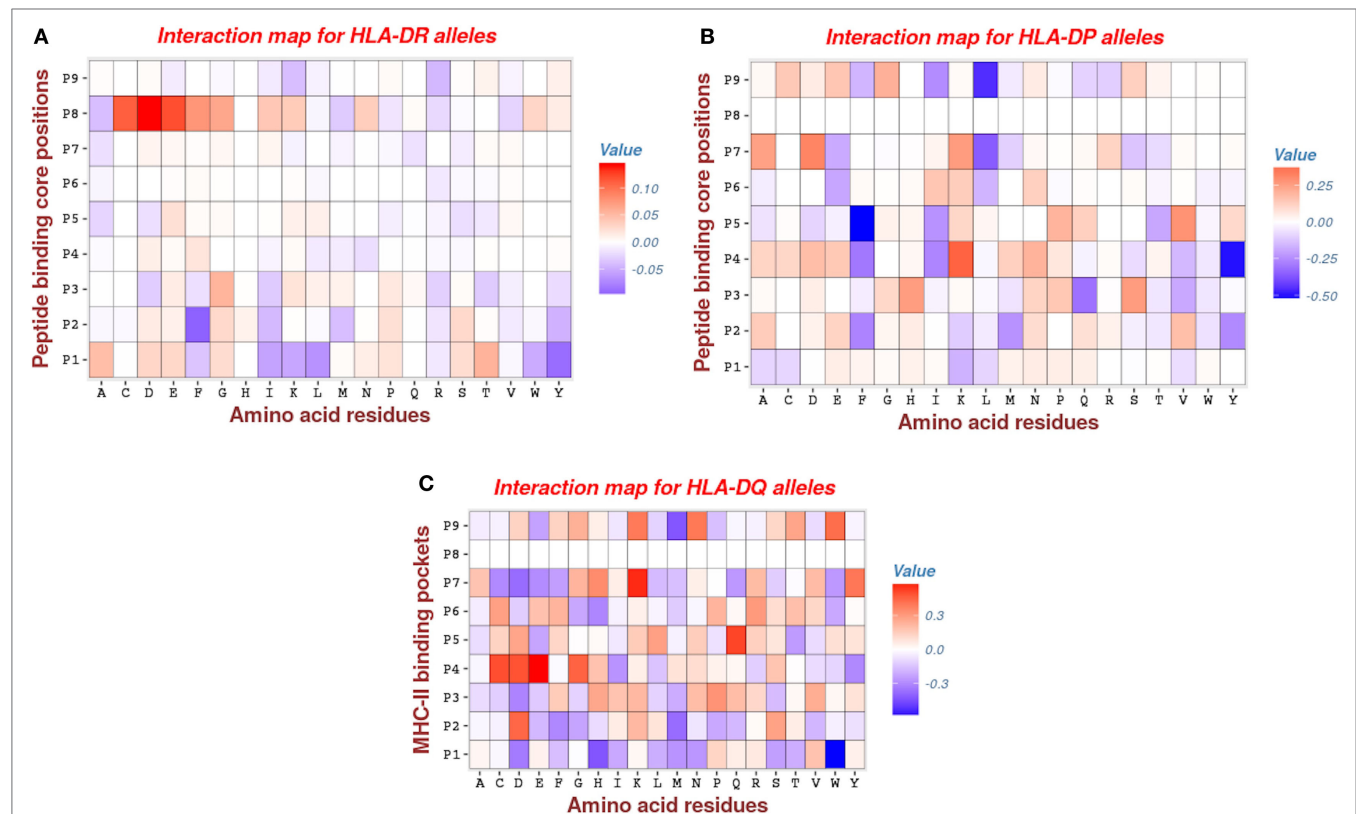
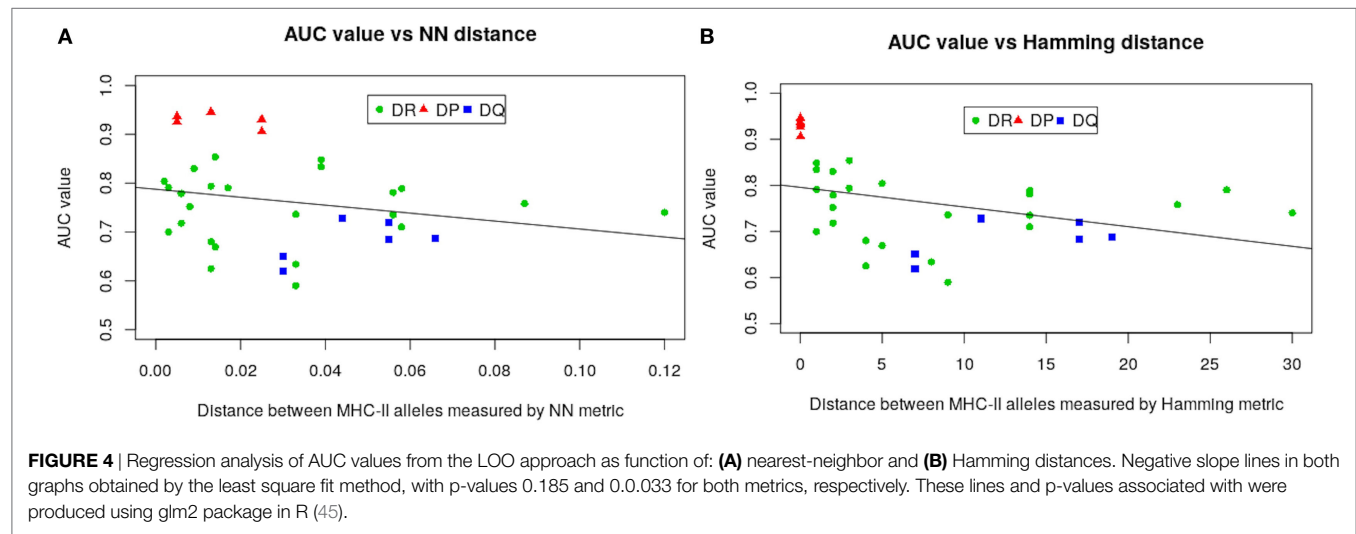


FIGURE 3 | Average performance results of the model in terms of AUC values for the three metrics: NN approach (gray bars), Hamming metric (blue bars), and the LOO method (red bars). Except for HLA-DQ loci, the LOO approach significantly out performs the other two metrics. Such results indicate that this method performs better than a random test even for un-characterized MHC-II molecules.

Analysis of HLA-DR parameters revealed that pocket *P1* has moderate attractive interactions with peptide (negative energies indicated by blue color in **Figure 5**), via hydrophobic (**I, L, W, Y**) side chains and, to lesser extent, via the aromatic (**F, W**) amino acids and a single hydrophilic residue (**K**). Remarkably, previous studies (2, 46) arrived at a similar conclusion of a large tendency of position *P1* toward interactions involving the hydrophobic

side chains. The repulsive interactions (positive energies indicated by red color in **Figure 5**) of pocket *P1* mostly occur with the hydrophilic side chains (**D, E, N, S, T**) and the aliphatic residue (**A**). Generally, most of the primary anchor pockets (*P1, P4, P6, P7, P9*) confer attractive interactions, but the pocket *P1* makes the largest contribution. This is consistent with results obtained using the MULTIRTA method (2). Among the secondary anchors,



we found that pocket *P2* has attractive interactions with aromatic (F, Y) and the hydrophobic (I, M, Y) side chains. The most repulsive interactions come from the pocket *P8*, which has a strong unfavorable interactions involving the side chains of residues C, D, E, F, G, I, L, W, and Y (see **Figure 5A**).

For HLA-DP, we found that pocket *P9* has significantly attractive interactions involving the hydrophobic residue (L). This is consistent with the previous results of Ref. (47) (see **Figure 5B**). Also, we found that pockets *P4* and *P5* have important attractive interactions with peptide via hydrophobic (Y) and aromatic (F) side chains, respectively. The contribution of the pocket *P4* is concordant with other studies such as (41), but the contribution of the pocket *P5* was not reported in the study of Andreatta and Nielsen (47), which was specifically dedicated to HLA-DQ and HLA-DP alleles. Furthermore, we found that the other two pockets *P1* and *P6*, which were reported as primary anchors in that study, have a moderate contribution to calculated bind energies (see **Figure 5B**).

The pattern of energetic contributions for HLA-DQ alleles is less ordered. There is no common pattern except the observation of significant attractive interaction of pocket *P1* via the hydrophobic residue (W) and the repulsive interaction of pocket *P4* via the side chains C, E, and D (see **Figure 5C**). This finding is in line with the observations of Morten et al. (47).

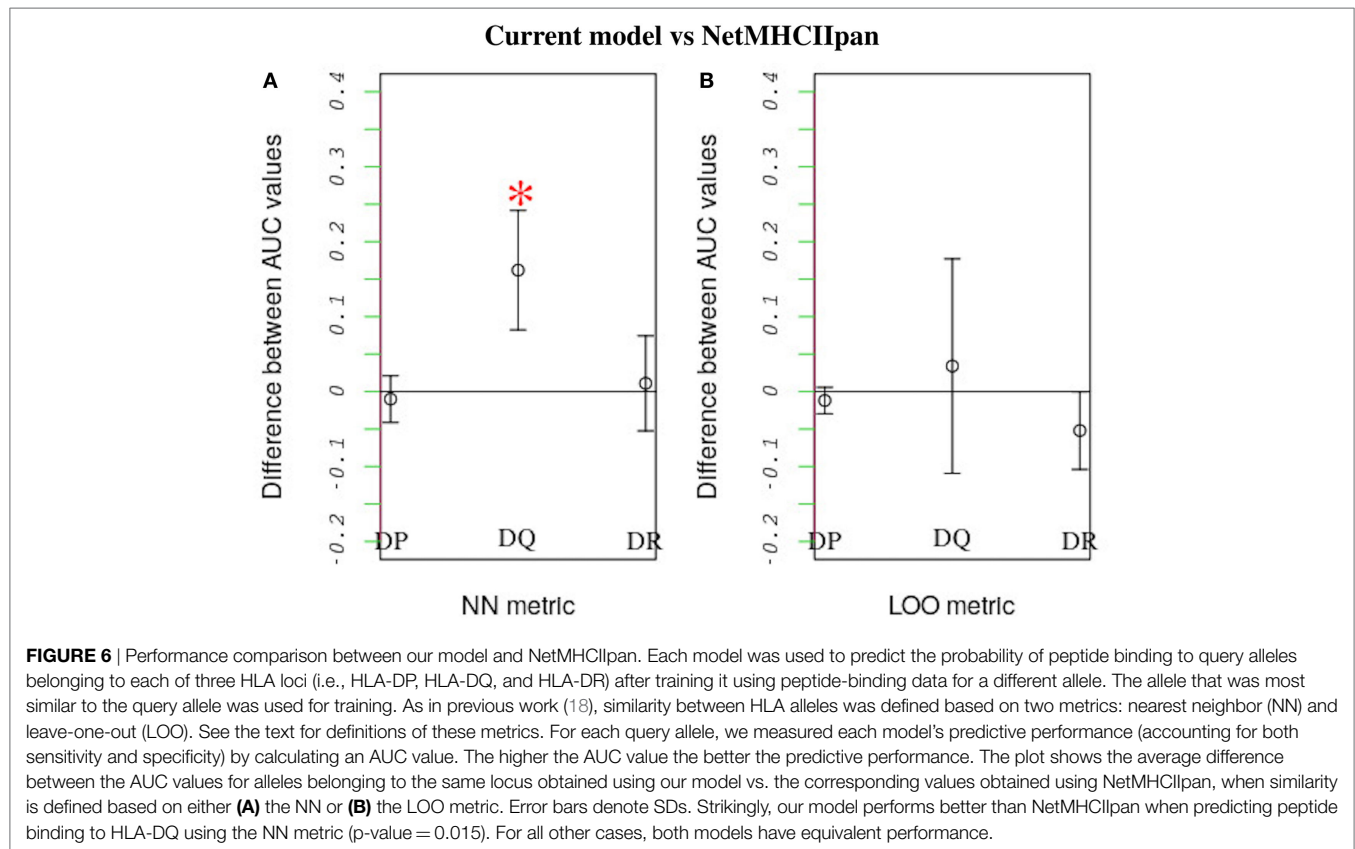
3.5. Discussion

Interactions between peptides and MHC-II molecules are central to the adaptive immune system. Precise prediction and

knowledge of the physicochemical determinants that govern such interaction is useful in designing effective and affordable epitope-based vaccines, and in providing insights about the immune system's mechanism as well as in understanding the pathogenesis of diseases. In this study, we have developed a trans-allelic model that can predict peptide interactions to the three human MHC-II loci. It can be readily applied to MHC-II molecules of other species provided that relative structural information are available. This method is based on biophysical ideas, an alternative to the dominant machine learning approaches.

The model presented here is, in addition to NetMHCIIpan, only the second trans-allelic method that allows comprehensive prediction analysis of peptide binding to all three human MHC-II loci. Most trans-allelic models for MHC-II peptides are restricted to HLA-DR and HLA-DP alleles. The TEPITOPEpan method (23), which is popular among immunologists and is the successor of a pioneer method in this field, is limited to HLA-DR alleles.

In this work we employed the definition of MHC polymorphic residue groups of the MULTIRTA method (2), which is more intuitive and inclusive than the MHC pseudo sequences of NetMHCIIpan (18), in developing our trans-allelic model. Utilizing new structural data for MHC-II complexes, which were not present when MULTIRTA was being developed, we extended that idea to cover all three human MHC-II loci. There exist similar exercises for capturing structural similarity among MHC molecules. The earlier works of Murthy and Stern (48)



and Sinigaglia and Hammer (49) were mostly limited to HLA-DR molecules. But in a previous study (2), the “polymorphic residue groups” were shown to be useful for inferring the interaction energy. This physical way of capturing structural similarity among MHC molecules works well in our biophysical approach.

We compared how well our model predicts the MHC-II allele binding preferences of a novel peptide dataset vs. how well the state-of-the-art NetMHCIIpan method performs the same task. In this comparison we applied both our model and NetMHCIIpan to predict binding preferences for peptides known to either bind or not bind a reference allele after training both models using peptide-binding data for a second allele. For a given MHC-II locus, the second allele was the one that was most similar to the reference allele. Similarity was quantified based on either a leave-one-out approach or a nearest-neighbor approach (see Section 3.3). When using the nearest-neighbor approach, we found that our model performs significantly better than NetMHCIIpan in predicting peptide-binding preferences for HLA-DQ alleles (P -value = 0.015; **Figure 6A**). Furthermore, at the 95% confidence level, for all other cases, we found no significant difference between the performances of the two models (**Figure 6**).

These results are reassuring and indicate that our inverse-physics approach constitutes a promising complement to the widely used pattern-based approach to peptide-MHC-II binding predictions. The outstanding predictive accuracy of the NetMHCIIpan is not the result of its theoretical basis. Rather it derives from the use of sophisticated ensembles of neural networks, which are very powerful. However, our method has a distinguishing advantage over all the advanced machine learning models in that it is more physically meaningful. It is worth noting that our prediction results of peptide-MHC-II interaction were based on *in silico* analysis of real data. Additional, *in vivo* and *in vitro* investigations are needed to further validate the reported predictive performance.

REFERENCES

- Janeway C, Travers P, Walport M, Shlomchik M. *Immunobiology: The Immune System in Health and Disease*. New York: Garland Science Publishing (2005).
- Bordner AJ, Mittelman HD. MultiRTA: a simple yet reliable method for predicting peptide binding affinities for multiple class II MHC allotypes. *BMC Bioinformatics* (2010) 11:482. doi:10.1186/1471-2105-11-482
- Lund O, Nielsen M, Lundegaard C, Kesmir C, Brunak S. *Immunological Bioinformatics*. Cambridge: MIT Press (2005).
- Kindt TJ, Goldsby RA, Osborne BA, Kuby J. *Kuby Immunology*. New York: W.H. Freeman (2007).
- Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. *Open Biol* (2013) 3:120139. doi:10.1098/rsob.120139
- Todd JA, Bell JL, Mcdevitt HO. LA-DQ β gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nat Biotechnol* (1987) 329(6140):599–604.
- Fogdell A, Hillert J, Sachs C, Olerup O. The multiple sclerosis- and narcolepsy-associated HLA class II haplotype includes the DRB5*0101 allele. *Tissue Antigens* (1995) 46(482):333–6. doi:10.1111/j.1399-0039.1995.tb02503.x
- Gough SCL, Simmonds MJ. The HLA region and autoimmune disease: associations and mechanisms of action. *Curr Genomics* (2007) 8(7):453–65. doi:10.2174/138920207783591690

DATA AVAILABILITY STATEMENT

The peptide dataset used to evaluate this method can be found in the [IEDB] (<http://tools.iedb.org/main/datasets/>), and the MHC-II sequences data also can be found in the [EMBL-EBI] (ftp://ftp.ebi.ac.uk/pub/databases/ipd/imgt/hla/fasta/DRB_prot.fasta).

AUTHOR CONTRIBUTIONS

All authors contributed equally to this work.

ACKNOWLEDGMENTS

The support of the DST-NRF Centre of Excellence in Mathematics and Statistical Sciences (CoE-MaSS) toward this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the CoE-MaSS. We also gratefully acknowledge the support of the Centre for High Performance Computing (CHPC) at Cape Town, South Africa, for providing us access to their computational facilities.

FUNDING

This work was supported by the following grants: AD is funded by DST-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS, award number BA2017/050) and the African Institute for Mathematical Sciences (AIMS) South Africa; WN is funded by the AIMS Global Secretariat.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <https://www.frontiersin.org/articles/10.3389/fimmu.2018.01410/full#supplementary-material>.

- Jones EY, Fugger L, Strominger JL, Siebold C. MHC class II proteins and disease: a structural perspective. *Nat Rev Immunol* (2006) 6(4):271–82. doi:10.1038/nri1805
- Mandic M, Castelli F, Janjic B, Almunia C, Andrade P, Gillet D, et al. One NY-ESO-1-derived epitope that promiscuously binds to multiple HLA-DR and HLA-DP4 molecules and stimulates autologous CD4+ T cells from patients with NY-ESO-1-expressing melanoma. *J Immunol* (2005) 174(3):1751–9. doi:10.4049/jimmunol.174.3.1751
- Soria-Guerra RE, Nieto-Gomez R, Govea-Alonso DO, Rosales-Mendoza S. An overview of bioinformatics tools for epitope prediction: implications on vaccine development. *J Biomed Inform* (2015) 53:405–14. doi:10.1016/j.jbi.2014.11.003
- Dimitrov I, Garnev P, Flower DR, Doytchinova I. MHC class II binding prediction—a little help from a friend. *J Biomed Biotechnol* (2010) 2010 705821. doi:10.1155/2010/705821
- Falk K, Rötzschke O, Stevanović S, Jung G, Rammensee HG. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* (1991) 351(6324):290–6. doi:10.1038/351290a0
- Wan J, Lin W, Xu Q, Ren Y, Flower DR, Li T. SVRMHC prediction server for MHC-binding peptides. *BMC Bioinformatics* (2006) 7:463. doi:10.1186/1471-2105-7-463
- Noguchi H, Kato R, Hanai T, Matsubara Y, Honda H, Brusica V, et al. Hidden Markov model-based prediction of antigenic peptides that interact with MHC

- class II molecules. *J Biosci Bioeng* (2002) 94(3):264–70. doi:10.1016/S1389-1723(02)80160-8
16. Nielsen M, Lundegaard C, Worning P, Lauemøller SL, Lamberth K, Buus S, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* (2003) 12(3):1007–17. doi:10.1110/ps.0239403
 17. Nielsen M, Lundegaard C, Blicher T, Peters B, Sette A, Justesen S, et al. Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS Comput Biol* (2008) 4(7):e1000107. doi:10.1371/journal.pcbi.1000107
 18. Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. *Immunogenetics* (2013) 65(10):711–24. doi:10.1007/s00251-013-0720-y
 19. Doytchinova I, Darren F. The HLA-A2-supermotif: a QSAR definition. *Org Biomol Chem* (2003) 7:2648–54. doi:10.1039/b300707c
 20. Bordner AJ, Mittelman HD. Prediction of the binding affinities of peptides to class II MHC using a regularized thermodynamic model. *BMC Bioinformatics* (2010) 11:41. doi:10.1186/1471-2105-11-41
 21. Zaitlen N, Reyes-Gomez M, Heckerman D, Jovic N. Shift-invariant adaptive double threading: learning MHC II-peptide binding. *J Comput Biol* (2008) 15(7):927–42. doi:10.1089/cmb.2007.0183
 22. Zhang L, Udaka K, Mamitsuka H, Zhu S. Toward more accurate pan-specific MHC-peptide binding prediction: a review of current methods and tools. *Brief Bioinform* (2012) 13(3):350–64. doi:10.1093/bib/bbr060
 23. Zhang L, Chen Y, Wong H-S, Zhou S, Mamitsuka H, Zhu S. TEPITOPEpan: extending TEPITOPE for peptide binding prediction covering over 700 HLA-DR molecules. *PLoS One* (2012) 7(2):e30483. doi:10.1371/journal.pone.0030483
 24. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* (2008) 61(1):1–13. doi:10.1007/s00251-008-0341-z
 25. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS One* (2007) 2(8):e796. doi:10.1371/journal.pone.0000796
 26. Jacob L, Vert J-P. Efficient peptide-MHC-I binding prediction for alleles with few known binders. *Bioinformatics* (2008) 24(3):358–66. doi:10.1093/bioinformatics/btm611
 27. Sturniolo T, Bono E, Ding J, Radrizzani L, Tuerci O, Sahin U, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol* (1999) 17(6):555–61. doi:10.1038/9858
 28. Nielsen M, Justesen S, Lund O, Lundegaard C, Buus S. NetMHCIIpan-2.0 – improved pan-specific HLA-DR predictions using a novel concurrent alignment and weight optimization training procedure. *Immunome Res* (2010) 6:9. doi:10.1186/1745-7580-6-9
 29. Andreatta M, Karosiene E, Rasmussen M, Stryhn A, Buus S, Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics* (2015) 67(11–12):641–50. doi:10.1007/s00251-015-0873-y
 30. Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* (2007) 8:238. doi:10.1186/1471-2105-8-238
 31. AlQuraishi M, Koytiger G, Jenney A, MacBeath G, Sorger PK. A multiscale statistical mechanical framework integrates biophysical and genomic data to assemble cancer networks. *Nat Genet* (2014) 46:1363–71. doi:10.1038/ng.3138
 32. Nguyen HC, Zecchina R, Berg J. Inverse statistical problems: from the inverse Ising problem to data science. *Adv Phys* (2017) 66(3):197–261. doi:10.1080/00018732.2017.1341604
 33. Holland CJ, Cole DK, Godkin A. Re-directing CD4(+) T cell responses with the flanking residues of MHC class II-bound peptides: the core is not enough. *Front Immunol* (2013) 4:172. doi:10.3389/fimmu.2013.00172
 34. Sette A, Adorini L, Colon SM, Buus S, Grey HM. Capacity of intact proteins to bind to MHC class II molecules. *J Immunol* (1989) 143(4):1265–7.
 35. Rose PW, Prlić A, Bi C, Bluhm WF, Christie CH, Dutta S, et al. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res* (2015) 43(Database issue):D345–56. doi:10.1093/nar/gku1214
 36. Li Z, Zhao Y, Pan G, Tang J, Guo F. A novel peptide binding prediction approach for HLA-DR molecule based on sequence and structural information. *Biomed Res Int* (2016) 2016:3832176. doi:10.1155/2016/3832176
 37. Bodenhofer U, Bonatesta E, Horejs-Kainrath C, Hochreiter S. msa: an R package for multiple sequence alignment. *Bioinformatics* (2015) 31(24):3997–9. doi:10.1093/bioinformatics/btv494
 38. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol* (1996) 58(1):267–88.
 39. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* (2010) 33(10):1–22. doi:10.18637/jss.v033.i01
 40. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* (2015) 43(Database issue):D405–12. doi:10.1093/nar/gku938
 41. Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* (2018). doi:10.1111/imm.12889
 42. Robinson J, Halliwell JA, McWilliam H, Lopez R, Marsh SGE. IPD—the immuno polymorphism database. *Nucleic Acids Res* (2013) 41(D1):D1234–40. doi:10.1093/nar/gks1140
 43. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* (2006) 27(8):861–874. doi:10.1016/j.patrec.2005.10.010
 44. Steven H, Jorja H. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* (1992) 89(22):10915–9. doi:10.1073/pnas.89.22.10915
 45. Marschner IC. glm2: fitting generalized linear models with convergence problems. *The R Journal* (2011) 3/2:12–15.
 46. Madden DR. The three-dimensional structure of peptide-MHC complexes. *Annu Rev Immunol* (1995) 13:587–622. doi:10.1146/annurev.13.040195.003103
 47. Andreatta M, Nielsen M. Characterizing the binding motifs of 11 common human HLA-DP and HLA-DQ molecules using NNAlign. *Immunology* (2012) 136(3):306–11. doi:10.1111/j.1365-2567.2012.03579.x
 48. Murthy VL, Stern LJ. The class II MHC protein HLA-DR1 in complex with an endogenous peptide: implications for the structural basis of the specificity of peptide binding. *Structure* (1997) 5(10):1385–96.
 49. Sinigaglia F, Hammer J. Rules for peptide binding to MHC class II molecules. *APMIS* (1994) 102(6):241–8. doi:10.1111/j.1699-0463.1994.tb04871.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Degoot, Chirove and Ndifon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Cross-Reactive Bactericidal Antimeningococcal Antibodies Can Be Isolated From Convalescing Invasive Meningococcal Disease Patients Using Reverse Vaccinology 2.0

Fadil A. Bidmos^{1*}, Simon Nadel^{1,2}, Gavin R. Screaton¹, J. Simon Kroll¹ and Paul R. Langford¹

¹ Section of Paediatrics, Department of Medicine, Imperial College London, London, United Kingdom, ² St. Mary's Hospital, Paddington, London, United Kingdom

OPEN ACCESS

Edited by:

Rino Rappuoli,
GlaxoSmithKline, Italy

Reviewed by:

Sanjay Ram,
University of Massachusetts
Medical School, United States
Mariagrazia Pizza,
GlaxoSmithKline, Italy

*Correspondence:

Fadil A. Bidmos
f.bidmos@imperial.ac.uk

Specialty section:

This article was submitted to
Vaccines and Molecular
Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 17 April 2018

Accepted: 29 June 2018

Published: 16 July 2018

Citation:

Bidmos FA, Nadel S, Screaton GR,
Kroll JS and Langford PR (2018)
Cross-Reactive Bactericidal
Antimeningococcal Antibodies Can
Be Isolated From Convalescing
Invasive Meningococcal Disease
Patients Using Reverse
Vaccinology 2.0.
Front. Immunol. 9:1621.
doi: 10.3389/fimmu.2018.01621

The threat from invasive meningococcal disease (IMD) remains a serious source of concern despite the licensure and availability of vaccines. A limitation of current serogroup B vaccines is the breadth of coverage afforded, resulting from the capacity for extensive variation of the meningococcus and its huge potential for the generation of further diversity. Thus, the continuous search for candidate antigens that will compose supplementary or replacement vaccines is mandated. Here, we describe successful efforts to utilize the reverse vaccinology 2.0 approach to identify novel functional meningococcal antigens. In this study, eight broadly cross-reactive sequence-specific antimeningococcal human monoclonal antibodies (hmAbs) were cloned from 4 ml of blood taken from a 7-month-old sufferer of IMD. Three of these hmAbs possessed human complement-dependent bactericidal activity against meningococcal serogroup B strains of disparate PorA and 4CMenB antigen sequence types, strongly suggesting that the target(s) of these bactericidal hmAbs are not PorA (the immunodominant meningococcal antigen), factor-H binding protein, or other components of current meningococcal vaccines. Reactivity of the bactericidal hmAbs was confirmed to a single ca. 35 kDa protein in western blots. Unequivocal identification of this antigen is currently ongoing. Collectively, our results provide proof-of-principle for the use of reverse vaccinology 2.0 as a powerful tool in the search for alternative meningococcal vaccine candidate antigens.

Keywords: *Neisseria meningitidis*, invasive meningococcal disease, reverse vaccinology 2.0, human monoclonal antibodies, vaccines

INTRODUCTION

Neisseria meningitidis is a major obligate human pathogen that frequently colonizes the nasopharynx asymptomatically, in a state known as carriage (1). Occasionally, invasive meningococcal disease (IMD) occurs, through invasion of pharyngeal tissues, proliferation in blood (meningococcal septicemia), and crossing of the blood–brain barrier leading to meningitis (2, 3). More than 70,000

cases of IMD are reported annually worldwide with case fatality ratios between 5 and 15%, even with therapeutic intervention (4, 5). Debilitating neurological sequelae are common among survivors of IMD (6–8).

The use of currently available polysaccharide conjugate meningococcal vaccines has been effective against targeted serogroups (mainly serogroups A, C, W, and Y) within vaccinated populations (9, 10). For serogroup B strains, which account for more than 60% of IMD in the UK and Europe (11, 12), vaccine development has focused heavily on sub-capsular vaccine components owing to the unsuitability of the serogroup B capsule (13). One of these vaccines, 4CMenB (or Bexsero®), is a protein-based vaccine whose major components are the factor-H binding protein (fHbp) (variant 1.1), Neisserial heparin-binding antigen (NHBA variant 2), Neisserial adhesin A (NadA variant 3), and the detergent-extracted outer membrane vesicle component of the New Zealand epidemic strain (with PorA variant P1.4) (14). Like the polysaccharide conjugate vaccines, accruing data shows high effectiveness of 4CMenB (15). However, we are seeing a gradual recrudescence of carriage and disease to pre-vaccine levels through vaccine-driven strain replacement (16–19). In addition, there are concerns that the changing epidemiology of IMD (20–22) may lead to a significant reduction in the efficacy of the vaccines in the long term. These limitations, coupled with the huge potential of the meningococcus to generate extensive antigenic diversity (leading to vaccine/immune escape) (23) justify the search for novel vaccine candidate antigens.

Preclinical vaccine development methods are enriched by detailed analysis of the human immune response to etiological agents of infectious diseases. For example, with the development of high-throughput technologies, deep sequencing of the gene segments encoding the variable regions of antibody heavy (V_H) and light ($V_L = \kappa$ or λ) chains in a given B cell repertoire is providing valuable information useful in understanding adaptive immunity to infections, autoimmunity, and malignancies (24, 25). Identifying the targets of antibodies of interest by cloning and *in vitro* expression of V_H and V_L chains of B-cell antibodies is a powerful approach, which can be utilized to inform on the functional immunogenicity of both known and novel antigens. The use of this approach, termed reverse vaccinology 2.0 (26), in the cloning of neutralizing human recombinant monoclonal antibodies [human monoclonal antibodies (hmAbs)] from patients convalescing from viral infectious diseases is well documented; the first studies in the use of reverse vaccinology 2.0 focused on the isolation and functional characterization of antibodies targeting the dengue, HIV, and influenza viruses (27–29). The power of the approach lies in the expression of paired V_H and V_L regions from individual plasmablasts or memory B cells; the output being the expression of hmAbs mimicking natural $V_H + V_L$ combinations induced in the host.

Because of the transience of peak plasmablast circulation [reviewed in Ref. (30)] and the higher incidence of IMD among infants and toddlers (placing a limitation on blood sample volume), we aimed to assess whether reverse vaccinology 2.0 could be employed in the discovery of novel meningococcal antigens of vaccine potential. In this brief research report, we will outline findings relating to the following aims: (i) whether cross-reactive

antimeningococcal hmAbs targeting surface proteins could be cloned from patient samples; and (ii) if these hmAbs possessed bactericidal activity against a wide panel of strains, specifically those not covered by the protein-based meningococcal vaccines.

MATERIALS AND METHODS

Ethics Statement and Study Participants

Studies with human blood samples were approved by the London—Fulham Research Ethics Committee (Ref.: 11/LO/1982). Informed written consent was obtained from patients or their representatives in accordance with the Declaration of Helsinki. Patients were recruited following admission to the Imperial Healthcare (St. Mary's Hospital) Paediatric Intensive Care Unit (PICU), London, UK.

Patient SM-P02 was a 7-month-old baby who presented with fever, irritability, and reduced oral intake. Rapidly spreading petechial rash was detected a few hours after arrival at the hospital. Blood cultures confirmed meningococcal septicemia. Molecular typing of the isolate, M14-240312, revealed that it was a serogroup B strain (MenB). A 4 ml blood sample was collected 7 days post-admission to St. Mary's Hospital PICU, following confirmation of negative blood cultures.

Bacterial Strains

Assays performed with whole bacteria or cell lysates involved patient isolates M14-240312 and a 16-strain meningococcal panel reflecting the current genetic epidemiological prevalence in the UK (obtained from the UK Meningococcal Reference Unit, Public Health England, Manchester) (Table 1). This 16-strain panel includes 10 strains genotypically mismatched for all the major 4CMenB antigens. Typing information was obtained from the

TABLE 1 | MenB strains composing the 17-strain screening panel.

S/N	Isolate	PorA		FetA VR	MLST	4CMenB antigen type		
		VR1	VR2			Factor-H binding protein	NHBA	NadA ^a
1	M14-240312	12-1	13-1	5-5	41/44	2.19	2	–
2	MC58	7	16-2	1-5	32	1.1	3	110
3	M07-240646	22	9	5-12	269	2.19	17	–
4	M07-240657	5-1	10-4	5-5	213	3.294	18	–
5	M07-240669	17-1	23	5-2	41/44	2.19	31	–
6	M07-240679	19	15	ND ^b	269	2.19	17	–
7	M07-240680	22	14-6	ND	41/44	2.19	43	–
8	M07-240909	22	14	5-5	213	2.19	18	–
9	M08-240014	22	9	ND	269	2.68	17	–
10	M08-240107	22	9	ND	269	3.59	17	–
11	M08-240164	22	14	ND	213	3.45	18	–
12	M08-240276	22	14	5-5	213	3.45	18	–
13	M10-240474	19-1	15-11	3-9	269	1.15	21	–
14	M10-240476	22	14	5-5	213	1.110	18	–
15	M10-240480	7-2	4	1-5	41/44	1.4	2	–
16	M11-240016	7	16	3-3	32	1.62	3	100
17	M11-240123	5-1	2-2	5-8	11	2.22	20	–

^aDash (–) denotes absence of NadA in isolate.

^bND denotes no data, i.e. no typing information available for this antigen.

Neisseria MLST database hosted on <https://pubmlst.org/neisseria/>. Meningococci were routinely grown on 5% horse blood agar or in brain heart infusion broth supplemented with 5% Levinthal's.

Cell Sorting, cDNA Synthesis, and V_H/V_L Cloning

Peripheral blood mononuclear cells (PBMCs) were extracted from a 4 ml blood sample using a density gradient centrifugation method (Leucosep) and sorted singly into individual wells of 96-well plates containing catch buffer (10 mM Tris pH 8.0, 10 U RNasin) as previously described (31). PBMCs were incubated with a cocktail of anti-human monoclonal antibodies targeting CD3 (PerCP/Cy5.5), CD14 (PerCP/Cy5.5), CD19 (RPE), CD20 (PerCP/Cy5.5), CD27 (FITC), CD38 (APC), and CD56 (PerCP/Cy5.5) for 30 min on ice, in the dark. Stained PBMCs were analyzed using a BD FACSARIA III cell sorter. Plasmablasts were gated as follows: CD3⁻, CD14⁻, CD19⁺, CD20⁻, CD56⁻, CD27^{high}, and CD38^{high}. A single freeze-thaw cycle was used to lyse cells and release RNA. cDNA was synthesized from released RNA using the QIAGEN OneStep RT-PCR kit, as per manufacturer's protocols. A nested PCR was used to amplify the V_H and V_L regions; primers contained restriction endonuclease sites, which facilitated cloning into respective AbVec-IgH (*AgeI* + *Sall*), AbVec-Igκ (*AgeI* + *BsiWI*), or AbVec-Igλ (*AgeI* + *XhoI*) expression vectors.

Expression of hmAbs in Human Embryonic Kidney (HEK-293) Cells

HEK-293 cells were transiently transfected with cognate plasmid pairs using polyethyleneimine as transfection agent, as previously described (31). Culture supernatants were harvested after 72 h. HmAbs were purified from culture supernatants with single-use Protein G spin columns (Ab SpinTrap, GE Healthcare), as per manufacturer's instructions.

Assessment of hmAb Specificity

HmAb reactivity to surface-bound meningococcal antigens was assessed using indirect cell-based ELISAs, as previously

described (32). Briefly, inactivated bacterial whole cell suspensions were normalized to an OD₆₀₀ of ~0.5. Whole cells (100 μl of cell suspensions) or 1 μg of 4CMenB in Carbonate-Bicarbonate buffer (Sigma-Aldrich, cat. no. C3041) were transferred into designated wells of flat-bottom polystyrene 96-well plates and incubated at 4°C overnight. Wells were subsequently incubated with 200 μl of blocking buffer (PBS, 0.05% Tween-20, 1% BSA) for 1 h at room temperature prior to addition of 100 μl of an appropriate dilution of hmAbs or plasma IgG. Following a 1-h incubation, wells were washed thrice to remove unbound antibodies. A 1:2,000 dilution of an anti-IgG alkaline phosphatase conjugate (Sigma-Aldrich, cat. no. A9544) was used to probe wells for 1 h at room temperature. Wells were washed five times before addition of a phosphatase substrate (Bio-Rad, cat. no. 172-1063) as per the manufacturer's instructions. Signal detection was performed at OD₄₀₅ using a microplate reader.

Further investigations into hmAb reactivity were performed using western blotting, as described elsewhere (33).

Serum Bactericidal Activity (SBA) Assay

Functional activity of antimeningococcal IgG was determined using the standardized SBA assay (34). Briefly, 10³ CFU meningococci were incubated with IgG (hmAb or purified total plasma IgG) (final assay concentration of 25%) and exogenous human complement (25%) for 90 min in a humidified CO₂ incubator at 37°C, with gentle shaking. IgG that effected a reduction in meningococcal CFU by >50% after 90 min compared to negative controls (complement only; IgG only) was considered to possess SBA activity.

RESULTS

Induction of a Functional Immune Response in Patient SM-P02

Flow cytometric analysis showed that the circulating plasmablast population in patient blood was low, measured at 0.3% (Figure 1). To assess whether these plasmablasts represented a functional response to the meningococcal infection, total IgG purified from

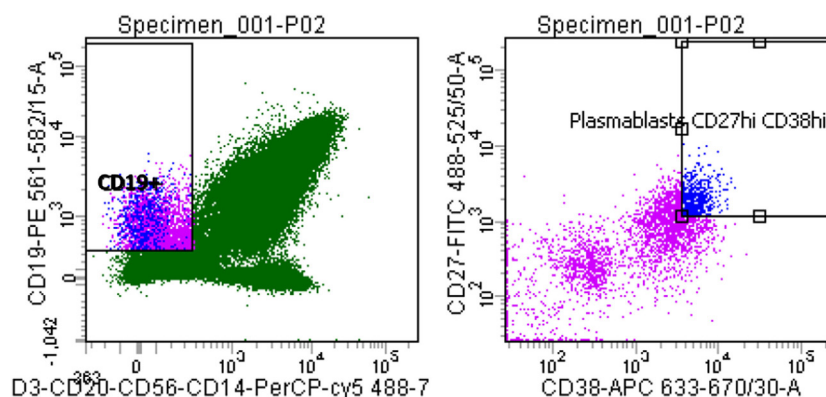


FIGURE 1 | Quantitative analysis of patient SM-P02's adaptive immune response to a MenB infection. Flow cytometry plots were gated for plasmablasts by CD3⁻, CD19⁺, then CD27^{high} and CD38^{high}.

patient plasma was assayed for bactericidal activity. Purified total IgG, rather than plasma, was assayed because of the presence of administered antibiotics in patient plasma. Reactivity of purified SM-P02 IgG to the patient isolate (M14-240312) and two other strains (MC58, which expresses the fHbp variant 1.1 in 4CMenB; and M08-240276, which is mismatched for all 4CMenB antigens) was assessed in ELISAs prior to assessment of SBA activity against the three strains. SM-P02 IgG bound to the infecting isolate, M14-240312, in cell-based ELISAs with corresponding SBA activity against the isolate. Despite significant binding of SM-P02 IgG to MenB strains MC58 and M08-240276 in ELISAs, no bactericidal activity was discerned against these strains. Thus, the induction of a specific bactericidal response in the patient to M14-240312 was confirmed (Figure 2A). Furthermore, the antimeningococcal immune response in patient SM-P02 was exclusive of the antigenic components of 4CMenB as no reactivity of SM-P02 plasma IgG to the vaccine antigens was discerned in ELISAs, under the experimental conditions employed in this study (Figure 2B).

In Vitro Cloning and Expression of Antimeningococcal Antibodies From Patient Plasmablasts

Amplification of the V_H and V_L (κ or λ) gene segments from 336 plasmablasts occurred at a PCR efficiency of 80%. Sequencing of these variable region gene segments showed biased usage of the IGHV3-7:IGLV2-8 gene pair, accounting for 16.7% of V gene usage. Of 139 recombinant hmAbs that were successfully expressed in HEK-293 cells, eight (Figure 3A) were reactive with the infecting MenB strain (M14-240312) and, in varying degrees, with the other members of the MenB strain panel (Figure 3B). None of these eight hmAbs were reactive with *Actinobacillus pleuropneumoniae* cells (a porcine respiratory pathogen). All antimeningococcal hmAbs possessed the IGHV3 class gene,

albeit with different subclasses and light chain gene pairs; the exception being P02-4F2 (IGHV4-59:IGKV1/1D-39). Two of these hmAbs (P02-5E10 and P02-6E9), however, possessed identical V gene pairs (IGHV3-30:IGKV4-1).

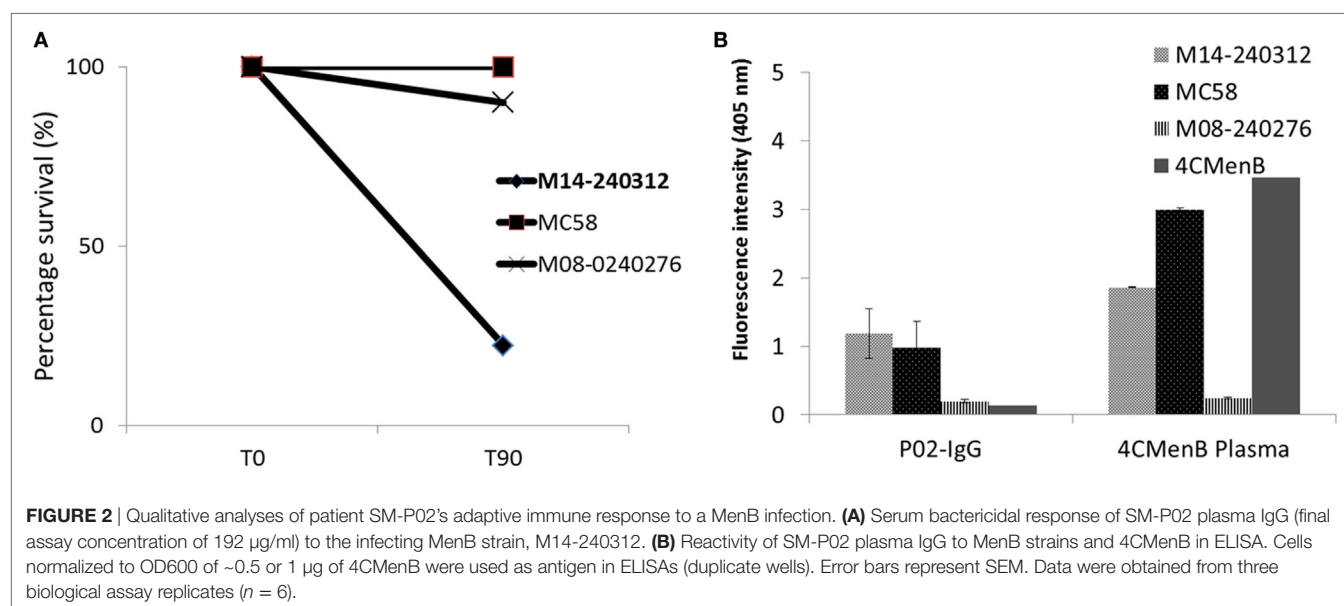
Specificity of the antimeningococcal hmAbs for the serogroup B capsule was ruled-out using the periodate assay, described in Ref. (35) (data not shown). Denaturing western blot data showed highly specific reactivity of three hmAbs (P02-1A1, P02-5E10, and P02-6E9) with a ~35 kDa meningococcal protein present in six members of the 17-strain panel (~35%). The target epitope of hmAb P02-1A1 was present in more strains ($n = 6$) than those of hmAbs P02-5E10 ($n = 3$) and P02-6E9 ($n = 2$) (Figure 4). The other five hmAbs, which were broadly cross-reactive with the strain panel (collectively recognizing 12 out of 17 strains), were non-reactive in denaturing western blots.

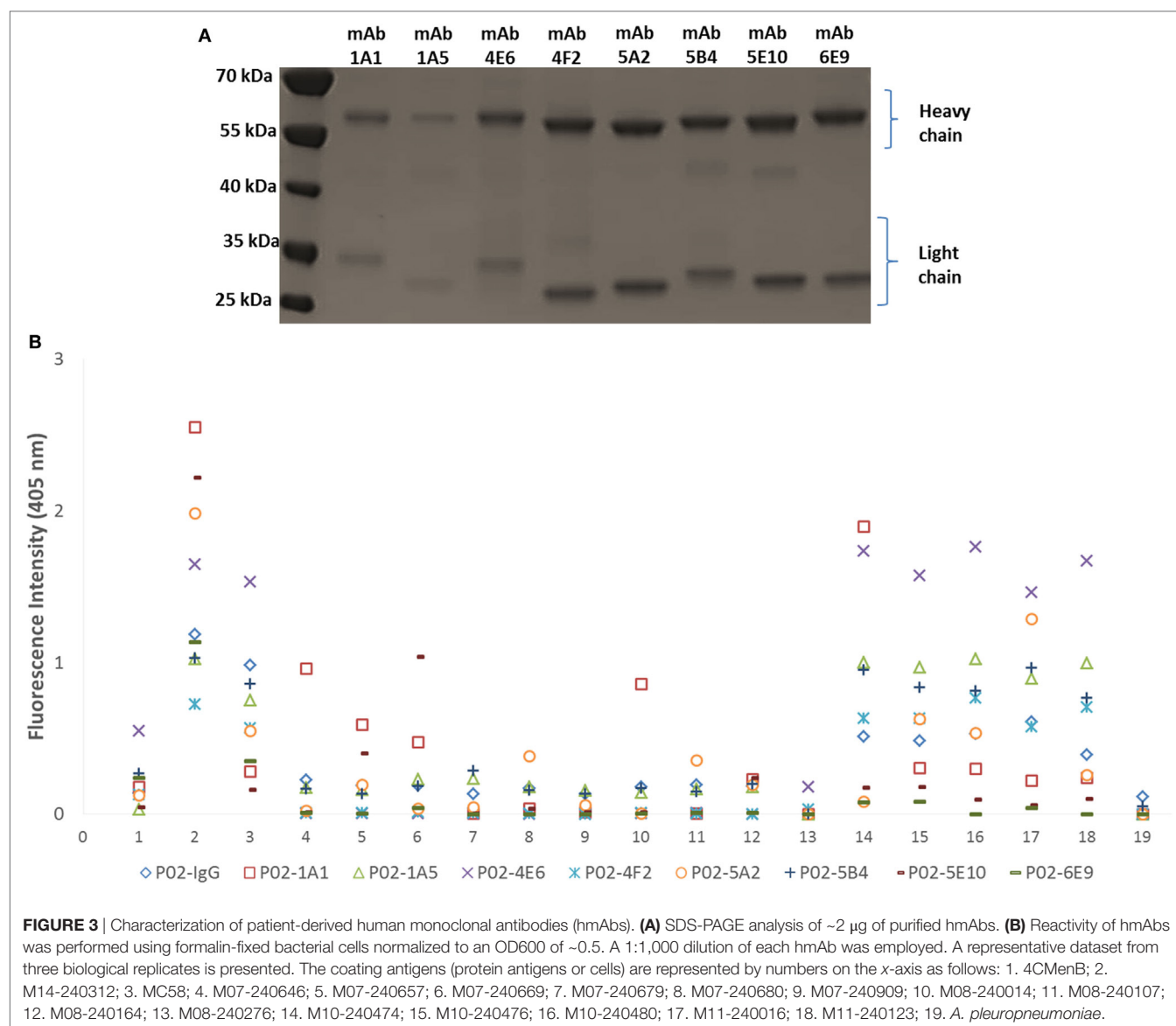
Cloned Antimeningococcal hmAbs Possess SBA Activity Against Patient Isolate M14-240312

To assess the functional activity of cloned recombinant hmAbs, SBAs were performed using purified hmAbs at a final assay concentration of 80 μ g/ml. Three hmAbs (P02-1A1, P02-5E10, and P02-6E9), in synergy with exogenous human complement, possessed SBA activity against strain M14-240312. All other hmAbs were non-bactericidal (Figure 5).

SBA Activity of hmAbs Is Strongly Linked to Surface Expression of Target Epitopes

To assess the breadth of SBA exhibited by hmAbs P02-1A1, P02-5E10, and P02-6E9, each hmAb was assayed for bactericidal activity against the 17-strain panel. HmAb P02-1A1-mediated killing of the patient isolate, M14-240312, and four other strains to which it bound in immunoassays (M07-240646, M07-240657, M08-240014, and M10-240474). Only one strain, M07-240669,





which showed positive reactivity with hmAb P02-1A1 in immunoassays was resistant to SBA activity of P02-1A1.

Consistent with results obtained with P02-1A1, hmAb P02-5E10 possessed SBA activity against strain M10-240474 (to which it bound in immunoassays) but not strain M07-240669, suggesting resistance of M07-240669 to complement-dependent killing. Interestingly, all three bactericidal hmAbs could mediate killing of other strains to which no discernible reactivity was found in immunoassays; hmAb P02-1A1 reproducibly mediated killing of strain M08-240276 while both P02-5E10 and P02-6E9 mediated killing of strain M10-240474 (Figure 6; Table 2).

DISCUSSION

Despite vaccination, the ongoing threat from IMD is unquestionable, especially in the African meningitis belt, justifying the search

for novel approaches to vaccine candidate discovery. Reverse vaccinology 2.0 has been useful in understanding the human adaptive immune response to disseminated viral (36, 37) and bacterial infections (38, 39). Hence, it presents as a potentially powerful tool that may identify novel meningococcal vaccine candidate antigens or reinforce the candidacy of some known antigens.

In the present study, we succeeded in isolating functional hmAbs from a 7-month-old patient convalescing from IMD despite challenging practical issues. The most affected age group for IMD is 6- to 24-months, and patients presenting at the St. Mary's PICU were mostly in this age group—children from whom the availability of sufficient quantities of blood sample for the isolation of PBMCs was very limited (4 ml). The isolation of specific antimeningococcal plasmablasts was further complicated by a paucity of knowledge on the magnitude

and timing of peak plasmablast responses following primary meningococcal infection in infants. The plasmablast population measured in the single sample obtained at 7 days post-admission analyzed in this study was 0.5%. Plasmablast population following acute infection with dengue virus (37) occurred at 6–7 days postinfection. While the induction of peak plasmablast response following infection with nosocomially acquired bacteria such as *Acinetobacter baumannii* occurred in most patients at 8–16 days with 40–80% of the total B-cell population being plasmablasts, they composed >2% of B-cell population in patient samples obtained at 0–7 days postinfection (38). The

low plasmablast induction seen in IMD patient SM-P02 may, therefore, be patient-specific and the timing of the peak level of cycling plasmablasts was likely missed (less than or more than 7 days post-admission). It is also plausible that low plasmablast induction is characteristic of a population at particular risk of IMD. Larger patient cohorts will be required to investigate this theory. Optimization of the IgG-cloning technique to increase the number of antimeningococcal plasmablasts using a recently published Ig capture-based assay (40) should improve productivity of the approach.

Reactivity of the antimeningococcal hmAbs isolated in this study with heterologous MenB strains possessing disparate PorA types strongly suggests that the antigen target of the hmAbs is not PorA, the immunodominant meningococcal antigen. It is likely that the three bactericidal hmAbs are reactive with a similar linear epitope contained in a yet-to-be identified ca. 35 kDa antigen. This 35 kDa antigen is surface-expressed in ~35% of the MenB strains, with heterologous 4CMenB antigen types, employed in this study. It is pertinent to note, however, that the ~35% presence of the linear epitope is specific to hmAbs generated in this study, and the antigen which it composes may possess more immunogenic surface-exposed epitopes and exhibit a wider presence among MenB and other meningococcal strains. One of these bactericidal hmAbs, P02-1A1, is reactive with a more diverse strain panel and possesses a higher bactericidal titer than hmAbs P02-5E10 and P02-6E9. This could be a result of somatic hypermutation of the P02-5E10 and P02-6E9 antibodies leading to the production of hmAb P02-1A1 with enhanced binding efficiency (affinity and/or avidity). Determination of the unequivocal identities of these hmAbs and their reactivities/SBA with other serogroups is currently ongoing. Taken together, data generated so far on these bactericidal hmAbs (P02-1A1, P02-5E10, and P02-6E9) strongly suggest that their target is not PorA, fHbp, or any of the 4CMenB recombinant antigens signifying its novelty and most importantly, candidacy for inclusion in future vaccine preparations. It is acknowledged, however,

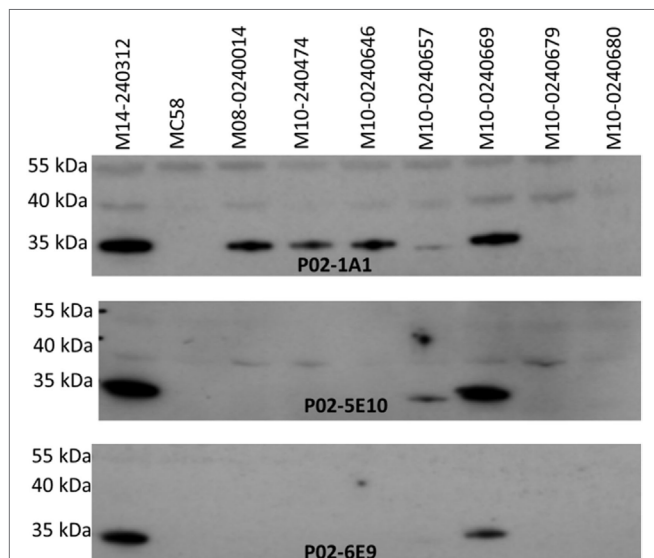


FIGURE 4 | Specificity of patient-derived antimeningococcal (hmAbs) for linear epitopes in denaturing western blots. Lysates from normalized suspensions of selected MenB strains were used as template in western blot experiments. A 1:1,000 dilution of each hmAb was employed.

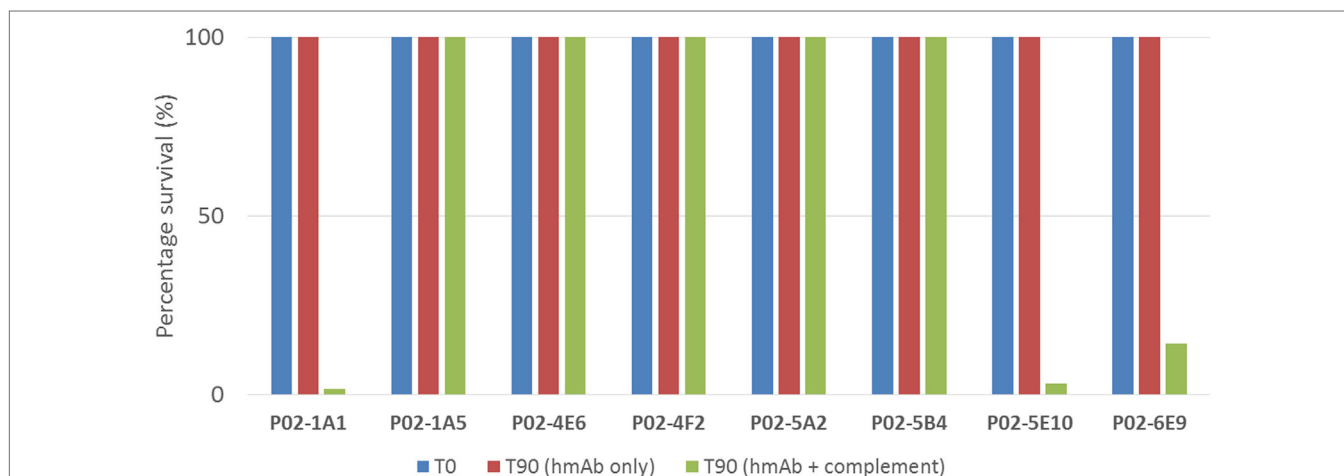


FIGURE 5 | SBA of patient-derived antimeningococcal human monoclonal antibodies (hmAbs) versus patient isolate, M14-240312. Assays were performed with hmAbs normalized to final assay concentration of 80 µg/ml, in three biological replicates. Percentage survival after 60-min incubation (T60) in the presence of 25% human complement versus inoculum (T0) is shown.

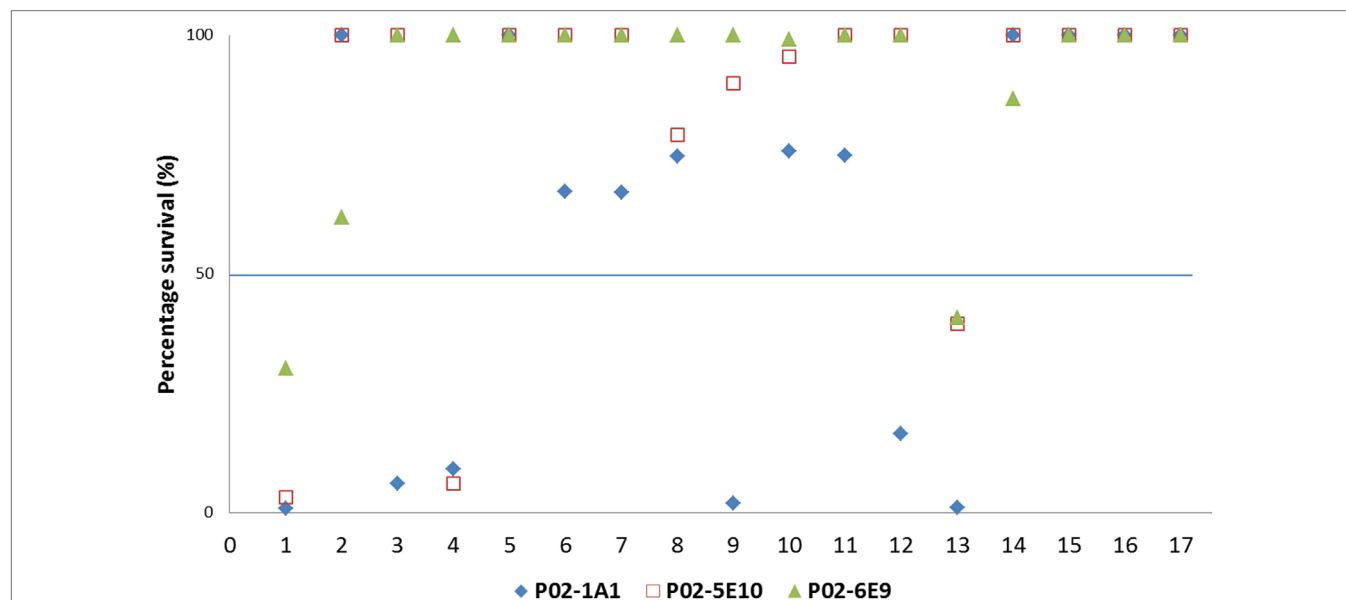


FIGURE 6 | SBA of P02-1A1, P02-5E10, and P02-6E9 versus MenB strains. Numbers represent strains as follows: 1. M14-240312; 2. MC58; 3. M07-0240646; 4. M07-0240657; 5. M07-0240669; 6. M07-0240679; 7. M07-0240680; 8. M07-0240909; 9. M08-0240014; 10. M08-0240107; 11. M08-0240164; 12. M08-0240276; 13. M10-240474; 14. M10-240476; 15. M10-240480; 16. M11-240016; 17. M11-240123. Assays were performed with hmAbs normalized to a final assay concentration of 80 µg/ml, in three biological replicates.

TABLE 2 | Summary table showing reactivity of the bactericidal human monoclonal antibodies—P02-1A1, P02-5E10, and P02-6E9—in immunoassays (ELISA and western blot), and SBA activity versus the 17-strain MenB panel.

S/N	Isolate	P02-1A1		P02-5E10		P02-6E9	
		I	SBA	I	SBA	I	SBA
1	M14-240312	+	+	+	+	+	+
2	MC58	–	–	–	–	+	–
3	M07-240646	+	+	–	–	–	–
4	M07-240657	+	+	+	+	–	–
5	M07-240669	+	–	+	–	+	–
6	M07-240679	–	–	–	–	–	–
7	M07-240680	–	–	–	–	–	–
8	M07-240909	–	–	–	–	–	–
9	M08-240014	+	+	–	–	–	–
10	M08-240107	–	–	–	–	–	–
11	M08-240164	–	–	–	–	–	–
12	M08-240276	–	+	–	–	–	–
13	M10-240474	+	+	–	+	–	+
14	M10-240476	–	–	–	–	–	–
15	M10-240480	–	–	–	–	–	–
16	M11-240016	–	–	–	–	–	–
17	M11-240123	–	–	–	–	–	–

I, immunoassay (ELISA and/or western blotting); SBA, serum bactericidal assay. Plus (+) denotes positive reactivity in immunoassays or bactericidal activity in SBA. Dash (–) denotes no discernible reactivity in immunoassays or bactericidal activity in SBA.

that while our current data suggest that the 35 kDa antigen is absent from the NZ OMV component of 4CMenB, further work is required to determine its unequivocal absence. Given the need for protein antigens that would compose improved or entirely novel cross-serogroup antimeningococcal vaccines, data from

this study show that reverse vaccinology 2.0 can be employed as a useful tool in identifying functionally immunogenic anti-meningococcal antigens.

ETHICS STATEMENT

Studies with human blood samples were approved by the London—Fulham Research Ethics Committee (Ref.: 11/LO/1982). Informed written consent was obtained from patients or their representatives in accordance with the Declaration of Helsinki. Patients were recruited following admission to the Imperial Healthcare (St. Mary's Hospital) Paediatric Intensive Care Unit (PICU), London, UK.

AUTHOR CONTRIBUTIONS

Conceptualization: PL, JK, SN, and GS; investigation: FB and PL; writing—original draft: FB; writing—reviewing and editing: FB, PL, SK, SN, and GS.

ACKNOWLEDGMENTS

We thank Stuart Gormley, Sobia Mustafa, and Michael Levin for their help in obtaining the patient sample, Hedda Wardemann for permitting the use of the AbVec vectors, and the Mongkolsapaya group for their help with the hmAb cloning protocol. This work was funded by a research grant from John and Michelle Bresnahan *via* MeningitisNow (awarded to PL, JK, SN, and GS) and an Imperial College Confidence-in-Concept Award (PL and FB).

REFERENCES

1. Caugant DA. Genetics and evolution of *Neisseria meningitidis*: importance for the epidemiology of meningococcal disease. *Infect Genet Evol* (2008) 8(5):558–65. doi:10.1016/j.meegid.2008.04.002
2. Thompson MJ, Ninin N, Perera R, Mayon-White R, Phillips C, Bailey L, et al. Clinical recognition of meningococcal disease in children and adolescents. *Lancet* (2006) 367(9508):397–403. doi:10.1016/S0140-6736(06)68609-1
3. van de Beek D, de Gans J, Tunkel AR, Wijdicks EF. Community-acquired bacterial meningitis in adults. *N Engl J Med* (2006) 354(1):44–53. doi:10.1056/NEJMra052116
4. Borrow R, Alarcón P, Carlos J, Caugant DA, Christensen H, Debbag R, et al. The global meningococcal initiative: global epidemiology, the impact of vaccines on meningococcal disease and the importance of herd protection. *Expert Rev Vaccines* (2017) 16(4):313–28. doi:10.1080/14760584.2017.1258308
5. Harrison LH, Trotter CL, Ramsay ME. Global epidemiology of meningococcal disease. *Vaccine* (2009) 27:B51–63. doi:10.1016/j.vaccine.2009.04.063
6. Chandran A, Herbert H, Misurski D, Santosham M. Long-term sequelae of childhood bacterial meningitis: an underappreciated problem. *Pediatr Infect Dis J* (2011) 30(1):3–6. doi:10.1097/INF.0b013e3181ef25f7
7. Edwards MS, Baker CJ. Complications and sequelae of meningococcal infections in children. *J Pediatr* (1981) 99(4):540–5. doi:10.1016/S0022-3476(81)80250-8
8. Kirsch EA, Barton RP, Kitchen L, Giroir BP. Pathophysiology, treatment and outcome of meningococemia: a review and recent experience. *Pediatr Infect Dis J* (1996) 15(11):967–79. doi:10.1097/00006454-199611000-00009
9. Kristiansen PA, Diomandé F, Ba AK, Sanou I, Ouédraogo AS, Ouédraogo R, et al. Impact of the serogroup A meningococcal conjugate vaccine, MenAfriVac, on carriage and herd immunity. *Clin Infect Dis* (2012) 56(3):354–63. doi:10.1093/cid/cis892
10. Trotter CL, Andrews NJ, Kaczmarski EB, Miller E, Ramsay ME. Effectiveness of meningococcal serogroup C conjugate vaccine 4 years after introduction. *Lancet* (2004) 364(9431):365–7. doi:10.1016/S0140-6736(04)16725-1
11. Villena R, Safadi MAP, Valenzuela MT, Torres JP, Finn A, O’Ryan M. Global epidemiology of serogroup B meningococcal disease and opportunities for prevention with novel recombinant protein vaccines. *Hum Vaccin Immunother* (2018) 14(5):1042–57. doi:10.1080/21645515.2018.1458175
12. Edge C, Waight P, Ribeiro S, Borrow R, Ramsay M, Ladhani S. Clinical diagnoses and outcomes of 4619 hospitalised cases of laboratory-confirmed invasive meningococcal disease in England: linkage analysis of multiple national databases. *J Infect* (2016) 73(5):427–36. doi:10.1016/j.jinf.2016.07.016
13. Finne J, Leinonen M, Mäkelä PH. Antigenic similarities between brain components and bacteria causing meningitis: implications for vaccine development and pathogenesis. *Lancet* (1983) 322(8346):355–7. doi:10.1016/S0140-6736(83)90340-9
14. Serruto D, Bottomley MJ, Ram S, Giuliani MM, Rappuoli R. The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: immunological, functional and structural characterization of the antigens. *Vaccine* (2012) 30:B87–97. doi:10.1016/j.vaccine.2012.01.033
15. Parikh SR, Andrews NJ, Beebejaun K, Campbell H, Ribeiro S, Ward C, et al. Effectiveness and impact of a reduced infant schedule of 4CMenB vaccine against group B meningococcal disease in England: a national observational cohort study. *Lancet* (2016) 388(10061):2775–82. doi:10.1016/S0140-6736(16)31921-3
16. Delrieu I, Yaro S, Tamekloe TA, Njanpop-Lafourcade BM, Tall H, Jaillard P, et al. Emergence of epidemic *Neisseria meningitidis* serogroup X meningitis in Togo and Burkina Faso. *PLoS One* (2011) 6(5):e19513. doi:10.1371/journal.pone.0019513
17. Soeters HM, Yameogo I, Sawadogo G, Aké F, Lingani C, Wang X, et al. Bacterial meningitis epidemiology and return of *Neisseria meningitidis* serogroup A cases in Burkina Faso in the five years following MenAfriVac mass vaccination campaign. *PLoS One* (2017) 12(11):e0187466. doi:10.1371/journal.pone.0187466
18. Sidikou F, Zaneidou M, Alkassoum I, Schwartz S, Issaka B, Obama R, et al. Emergence of epidemic *Neisseria meningitidis* serogroup C in Niger, 2015: an analysis of national surveillance data. *Lancet Infect Dis* (2016) 16(11):1288–94. doi:10.1016/S1473-3099(16)30253-5
19. Trotter CL, Lingani C, Fernandez K, Cooper LV, Bitá A, Tevi-Benissan C, et al. Impact of MenAfriVac in nine countries of the African meningitis belt, 2010–15: an analysis of surveillance data. *Lancet Infect Dis* (2017) 17(8):867–72. doi:10.1016/S1473-3099(17)30301-8
20. Gray SJ, Trotter CL, Ramsay ME, Guiver M, Fox AJ, Borrow R, et al. Epidemiology of meningococcal disease in England and Wales 1993/94 to 2003/04: contribution and experiences of the Meningococcal Reference Unit. *J Med Microbiol* (2006) 55(7):887–96. doi:10.1099/jmm.0.46288-0
21. Halperin SA, Bettinger JA, Greenwood B, Harrison LH, Jelfs J, Ladhani SN, et al. The changing and dynamic epidemiology of meningococcal disease. *Vaccine* (2012) 30:B26–36. doi:10.1016/j.vaccine.2011.12.032
22. Whittaker R, Dias JG, Ramliden M, Ködmön C, Economopoulou A, Beer N, et al. The epidemiology of invasive meningococcal disease in EU/EEA countries, 2004–2014. *Vaccine* (2017) 35(16):2034–41. doi:10.1016/j.vaccine.2017.03.007
23. Mowlaboccus S, Perkins TT, Smith H, Sloots T, Tozer S, Premph LJ, et al. Temporal changes in BEXSERO® antigen sequence type associated with genetic lineages of *Neisseria meningitidis* over a 15-year period in Western Australia. *PLoS One* (2016) 11(6):e0158315. doi:10.1371/journal.pone.0158315
24. DeKosky BJ, Ippolito GC, Deschner RP, Lavinder JJ, Wine Y, Rawlings BM, et al. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol* (2013) 31(2):166. doi:10.1038/nbt.2492
25. Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol* (2014) 32(2):158. doi:10.1038/nbt.2782
26. Rappuoli R, Bottomley MJ, D’Oro U, Finco O, De Gregorio E. Reverse vaccinology 2.0: human immunology instructs vaccine antigen design. *J Exp Med* (2016) 213(4):469–81. doi:10.1084/jem.20151960
27. Schieffelin JS, Costin JM, Nicholson CO, Orgeron NM, Fontaine KA, Isern S, et al. Neutralizing and non-neutralizing monoclonal antibodies against dengue virus E protein derived from a naturally infected patient. *Virol J* (2010) 7(1):28. doi:10.1186/1743-422X-7-28
28. Wrammert J, Smith K, Miller J, Langley WA, Kokko K, Larsen C, et al. Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* (2008) 453(7195):667. doi:10.1038/nature06890
29. Wu X, Yang ZY, Li Y, Hogerkorp CM, Schief WR, Seaman MS, et al. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* (2010) 329(5993):856–61. doi:10.1126/science.1187659
30. Fink K. Origin and function of circulating plasmablasts during acute viral infections. *Front Immunol* (2012) 3:78. doi:10.3389/fimmu.2012.00078
31. Smith K, Garman L, Wrammert J, Zheng NY, Capra JD, Ahmed R, et al. Rapid generation of fully human monoclonal antibodies specific to a vaccinating antigen. *Nat Protoc* (2009) 4(3):372. doi:10.1038/nprot.2009.3
32. Kohl TO, Ascoli CA. Direct and indirect cell-based enzyme-linked immunosorbent assay. *Cold Spring Harb Protoc* (2017) 2017(5):402–7. doi:10.1101/pdb.prot093757
33. Bidmos FA, Chan H, Praekelt U, Tauseef I, Ali YM, Kaczmarski EB, et al. Investigation into the antigenic properties and contributions to growth in blood of the meningococcal haemoglobin receptors, HpuAB and HmbR. *PLoS One* (2015) 10(7):e0133855. doi:10.1371/journal.pone.0133855
34. Borrow R, Aaberge IS, Santos GE, Eudaly TL, Oster P, Glennie A, et al. Interlaboratory standardization of the measurement of serum bactericidal activity by using human complement against meningococcal serogroup B, strain 44/76-SL, before and after vaccination with the Norwegian MenBvac outer membrane vesicle vaccine. *Clin Diagn Lab Immunol* (2005) 12(8):970–6. doi:10.1128/CDLI.12.8.970-976.2005
35. Woodward MP, Young WW Jr, Bloodgood RA. Detection of monoclonal antibodies specific for carbohydrate epitopes using periodate oxidation. *J Immunol Methods* (1985) 78(1):143–53. doi:10.1016/0022-1759(85)90337-0
36. Goodwin E, Gilman MS, Wrapp D, Chen M, Ngwuta JO, Moin SM, et al. Infants infected with respiratory syncytial virus generate potent neutralizing antibodies that lack somatic hypermutation. *Immunity* (2018) 48(2):339–49. doi:10.1016/j.immuni.2018.01.005
37. Wrammert J, Onlamoon N, Akondy RS, Perng GC, Polsrila K, Chandele A, et al. Rapid and massive virus-specific plasmablast responses during acute dengue virus infection in humans. *J Virol* (2012) 86(6):2911–8. doi:10.1128/JVI.06075-11

38. Band VI, Ibegbu C, Kaur SP, Cagle SM, Tribble R, Jones CL, et al. Induction of human plasmablasts during infection with antibiotic-resistant nosocomial bacteria. *J Antimicrob Chemother* (2014) 69(7):1830–3. doi:10.1093/jac/dku047
39. Zimmermann N, Thormann V, Hu B, Köhler AB, Imai-Matsushima A, Loch C, et al. Human isotype-dependent inhibitory antibody responses against *Mycobacterium tuberculosis*. *EMBO Mol Med* (2016) 8(11):1325–39. doi:10.15252/emmm.201606330
40. Pinder CL, Kratochvil S, Cizmeci D, Muir L, Guo Y, Shattock RJ, et al. Isolation and characterization of antigen-specific plasmablasts using a novel flow cytometry-based Ig capture assay. *J Immunol* (2017) 199(12):4180–8. doi:10.4049/jimmunol.1701253

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MP and handling Editor declared their shared affiliation.

Copyright © 2018 Bidmos, Nadel, Screatton, Kroll and Langford. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Web Resource for Designing Subunit Vaccine Against Major Pathogenic Species of Bacteria

Gandharva Nagpal^{1,2†}, Salman Sadullah Usmani^{1,3†} and Gajendra P. S. Raghava^{1,3*}

¹ Bioinformatics Centre, CSIR-Institute of Microbial Technology, Chandigarh, India, ² Centre for Bioinformatics, Computational and Systems Biology, Pathfinder Research and Training Foundation, Greater Noida, India, ³ Center for Computational Biology, Indraprastha Institute of Information Technology, Okhla, India

OPEN ACCESS

Edited by:

Pedro A. Reche,
Complutense University of Madrid,
Spain

Reviewed by:

Giampiero Pietrocola,
University of Pavia, Italy
Ghita Ghislat,
INSERM U1104 Centre
d'immunologie de Marseille-Luminy,
France

*Correspondence:

Gajendra P. S. Raghava
raghava@iiitd.ac.in

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Vaccines and Molecular Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 29 June 2018

Accepted: 13 September 2018

Published: 02 October 2018

Citation:

Nagpal G, Usmani SS and
Raghava GPS (2018) A Web Resource
for Designing Subunit Vaccine Against
Major Pathogenic Species of Bacteria.
Front. Immunol. 9:2280.
doi: 10.3389/fimmu.2018.02280

Evolution has led to the expansion of survival strategies in pathogens including bacteria and emergence of drug resistant strains proved to be a major global threat. Vaccination is a promising strategy to protect human population. Reverse vaccinology is a more robust vaccine development approach especially with the availability of large-scale sequencing data and rapidly dropping cost of the techniques for acquiring such data from various organisms. The present study implements an immunoinformatic approach for screening the possible antigenic proteins among various pathogenic bacteria to systemically arrive at epitope-based vaccine candidates against 14 pathogenic bacteria. Thousand four hundred and fifty nine virulence factors and Five hundred and forty six products of essential genes were appraised as target proteins to predict potential epitopes with potential to stimulate different arms of the immune system. To address the self-tolerance, self-epitopes were identified by mapping on 1000 human proteome and were removed. Our analysis revealed that 21 proteins from 5 bacterial species were found as virulent as well as essential to their survival, proved to be most suitable vaccine target against these species. In addition to the prediction of MHC-II binders, B cell and T cell epitopes as well as adjuvants individually from proteins of all 14 bacterial species, a stringent criteria lead us to identify 252 unique epitopes, which are predicted to be T-cell epitopes, B-cell epitopes, MHC II binders and Vaccine Adjuvants. In order to provide service to scientific community, we developed a web server VacTarBac for designing of vaccines against above species of bacteria. This platform integrates a number of tools that includes visualization tools to present antigenicity/epitopes density on an antigenic sequence. These tools will help users to identify most promiscuous vaccine candidates in a pathogenic antigen. This server VacTarBac is available from URL (<http://webs.iiitd.edu.in/raghava/vactarbac/>).

Keywords: reverse vaccinology, vaccine designing, immunotherapeutic, epitopes, antigen, virulence factor, essential genes

INTRODUCTION

Evolution of existing bacterial pathogens and emergence of new pathogenic strains are continuously causing problems to mankind. To address the pathogenic challenges to the human health, researchers developed various vaccines and antibiotics during the twentieth century. A worldwide usage of such therapeutic strategies led to the expansion of structural as well as

behavioral properties in various bacterial species. Diverse defense mechanisms were acquired by the known bacterial species that resulted in antibiotic tolerance or drug resistance, currently a major concern during administration of known antibiotics in pathogenic conditions. Lord Jim O'Neil and his team estimated that antimicrobial resistance could cause 10 million deaths a year by 2050 (1).

Vaccination has proved to be a promising strategem to protect human population from dreadful diseases like smallpox, polio, etc. Yet, vaccines are currently unavailable for many infectious diseases such as melioidosis caused by *Burkholderia pseudomallei* (2). Even among the existing vaccines, BCG which is heavily used vaccine against tuberculosis, is less effective in the immunocompromised and adults in tropical and subtropical region (3) and its efficacy is rapidly decreasing (4). Anthrax, DPT, Hib, Meningococcal vaccine, Pneumococcus heptavalent conjugate polysaccharide vaccine (PCV7), Pneumococcus 13-valent conjugate polysaccharide vaccine (PCV13) etc are generally considered effective at controlling disease symptoms. However, a major concern of recent vaccination programs tends to be the preference of vaccine escape mutants, i.e., the repertoire of immune targets in the pathogen is different from that used in the vaccine formulation (5). Antigenic variation and continuous bacterial evolution has necessitated designing of better vaccine candidates.

Virulence factors (VFs) have been explored to design vaccines against several pathogens such as Porcine enteric coronaviruses (CoVs) (6), *Vibrio cholerae*, enterotoxigenic *Escherichia coli* (7) etc. Bacterial virulence factors (VFs) are captivating, as they play a major role in establishment of infection to cause pathological conditions while surviving in a hostile environment (8). Virulence factors are thus defined as the gene products enabling the pathogen to enter, replicate, and persist in a host even in a small inoculum. These include cell surface proteins that facilitate bacterial attachment, cell surface carbohydrates which protect a bacterium, bacterial toxins, and hydrolytic enzymes that subsidize pathogenicity of bacterium (8).

In addition to virulence factors, genes essential for the living of the bacterial cell, have proved to be attractive drug targets in various pathogens such as the fungus *Aspergillus fumigatus* (9, 10), protozoan *Leishmania* species, (11) as well as bacteria like *Streptococcus pneumoniae*, *Haemophilus influenza*, and *Moraxella catarrhalis* (11). Essential genes are indispensable for cellular metabolism.

The last two decades witnessed an embracement of a more revolutionary vaccine development approach, namely the information driven “reverse vaccinology” which involves identification of vaccine candidates by genome or proteome sequence analysis as the primary step. High cost and practical limitations of feasibility and time have engendered a predilection toward computational techniques for rational reduction in the number of vaccine candidates to be experimentally tested. The epitope prediction methods have made possible searching of sequence sets as large as the whole proteomes of organisms with various restriction filters to arrive at novel vaccine candidates.

In order to facilitate the scientific community, researchers have developed numerous pipelines, databases, and tools for

identifying vaccine candidates against specific types of diseases (12, 13). Recently, researchers have proposed potential vaccine candidates against diverse pathogens like *Ebola* (14), *Zika* (15), *M. tuberculosis* (16), *Shigella* (17), *Hepatitis B Virus* (18), *Helicobacter pylori* (19), *Vibrio cholera* (20), *Dengue* (21), etc. using *in silico* predictions. These predicted candidates help reduce the burden in experimental labs for the researchers, thus saving time and cost of vaccine development. Most of the above-mentioned studies only focus on a single pathogen and lacked identification of adjuvant candidates.

In this study, we have systemically arrived at epitope based vaccine candidate against major pathogenic bacterial species. **Table 1** provides a summary of the major pathogenic manifestations of the bacterial species considered in the present study. Both, the virulence factors and the essential genes, have been considered to predict epitopes. It is a well-known fact that, diverse bacterial species and strains have evolved multiple mechanisms for evading the host immune responses, not only by producing likes of host inhibition pathway receptor ligands (“molecular mimicry”) but also by producing virulence factors resembling the intermediates of the host inhibitory pathways. Hence, peptides found in proteomes of the 1000 human genomes were removed while implementing the pipeline. To assist the scientific community all the potential vaccine candidates have been compiled and presented in the form of a database (<http://webs.iitd.edu.in/raghava/vactarbac/>).

MATERIALS AND METHODS

Raw Data Source

Proteins were taken as raw data to predict the potential vaccine candidates. We considered both virulence factors as well as essential genes of pathogenic bacteria.

Virulence Factors

The Virulence Factor Database (VFDB) is a resource for housing the Virulence Factors (VFs) that enable a microorganism to establish itself on or within a host thus bestowing the pathogen with a potential to cause disease (22). VFDB provides two hierarchical non-redundant datasets, i.e., sets A and B. Set A is a core dataset, covering genes associated with experimentally verified virulence factors. Therefore, we have selected protein sequences of core dataset A for further analysis. The link http://www.mgc.ac.cn/VFs/Down/VFDB_setA_pro.fas.gz provides protein sequences of the VFDB database in FASTA format when uncompressed using the command “gunzip” on a unix command line or using the tool “Winzip” on a Windows platform.

Essential Genes/ Proteins

Database of Essential Genes (DEG) maintains currently known essential genomic elements, such as protein-coding genes and non-coding RNAs, among bacteria, archaea, and eukaryotes (22). These genes are mandatory for survival of the corresponding bacterium. DEG contains 18286 proteins encoded from genes experimentally found to be essential for the survival of the 32 bacteria. For our study, we restricted the analysis of bacterial

TABLE 1 | Bacterial species considered in the study with the pathological conditions caused by them.

S. No.	Bacterial Species	Human Disease
1	<i>Bacillus subtilis</i>	food poisoning
2	<i>Burkholderia pseudomallei</i>	Melioidosis
3	<i>Campylobacter jejuni</i>	Campylobacteriosis
4	<i>Escherichia coli</i>	cholecystitis, bacteremia, cholangitis, urinary tract infection (UTI), and traveler's diarrhea, and other clinical infections such as neonatal meningitis and pneumonia
5	<i>Haemophilus influenzae</i>	pneumonia, bacteremia, meningitis, epiglottitis, septic arthritis, cellulitis, otitis media, and purulent pericarditis
6	<i>Helicobacter pylori</i>	ulcers in the stomach and small intestine
7	<i>Mycobacterium tuberculosis</i>	Tuberculosis
8	<i>Pseudomonas aeruginosa</i>	urinary tract infections, respiratory system infections, dermatitis, soft tissue infections, bacteremia, bone and joint infections, gastrointestinal infection
9	<i>Salmonella enteric</i>	Food poisoning
10	<i>Staphylococcus aureus</i>	pneumonia, meningitis, osteomyelitis, endocarditis, toxic shock syndrome, bacteremia, and sepsis
11	<i>Streptococcus agalactiae</i>	neonatal infection, septicemia, pneumonia, meningitis, chorioamnionitis
12	<i>Streptococcus pneumoniae</i>	pneumonia (infection of the lungs), ear infections, sinus infections, meningitis (infection of the covering around the brain and spinal cord), and bacteremia (blood stream infection)
13	<i>Streptococcus pyogenes</i>	streptococcal pharyngitis, rheumatic fever, rheumatic heart disease, and scarlet fever
14	<i>Vibrio cholera</i>	Cholera

essential genes belonging to 14 species that have also been included in VFDB, for the convenience of comparison and combining with VFDB analysis.

Target Selection for Vaccine Identification

The most important aspect of subunit vaccine designing is vaccine target selection. In case of data downloaded from VFDB (set A), it contains 2595 proteins (11) from 46 bacterial species. We have matched these bacterial species into DEG database that resulted into 14 common bacterial species. Therefore, our dataset contains only those bacterial species, which have at least one virulence factor and one protein encoded by essential genes from VFDB and DEG respectively. DEG contains 9664 proteins corresponding to these pathogenic bacteria. These are the essential proteins of pathogenic bacteria. Some specific bacterial components, because of their function proved to be better possible candidate for designing subunit vaccine (20, 23–25) therefore; we have selected some specific significant proteins.

Membrane Protein

These are the protein emerged from the membrane of various pathogenic bacteria. There are evidences that membrane proteins

mediate pathogen entry and colonization and are also the primary target of adaptive immune response (24). A total of 420 membranous proteins were selected from DEG for further processing.

Envelope Protein

These proteins are derived from the envelope of various pathogenic bacteria and are essential for their survival. These are important for attachment to host cell surface and also responsible for facilitating an immune response in the host cell (25). A total of 36 proteins were screened for further processing.

Secretion Protein

Fifty-two secretory proteins corresponding to bacterial species of our interest were deposited in DEG and play important role in their survival. Secretory system of pathogenic bacteria has been heavily exploited for screening of new possible vaccine (26, 27). These proteins play many roles in promoting bacterial virulence, from enhancing attachment to eukaryotic cells, to scavenging resources in an environmental niche, to directly intoxicating target cells and disrupting their functions (28).

Repair Protein

Bacterial damage repair mechanisms have broader roles encompassing responses to virulence as well as stress (29). Therefore we have also considered 38 proteins associated with the repair mechanism of various pathogenic bacteria.

Epitope Prediction Pipeline

Generation of Nona-Peptides

In order to design new vaccine candidate, first we have generated nona-peptides. Nona-peptides were 9-mer sequences (9 residues continuous stretch of peptide) originated from the essential or virulent proteins selected for the study. In order to avoid redundancy, we have removed all duplicated nona-peptides.

Removal of Self-Epitopes

Self-tolerance must also be considered in vaccine designing, as body's immune system rarely acts against self-epitopes. Therefore, all the nona-peptides, which are also present in human body need to be removed. To achieve this, we mapped nona-peptides on 1000 human proteome, and removed all the nona-peptides which are 100% identical.

Epitope Prediction

Our subsequent objective was to select peptides, which could activate human immune system, and generate memory cells. Therefore, we have used a pipeline for predicting different kinds of immune epitopes among these nona-peptides. This pipeline predicts; (i) B-cell epitopes using LBtope (30) (ii) MHC class II binders using ProPred (31) (iii) T-cell epitopes using CTLpred (32) and (iv) vaccine adjuvant using VaxinPAD (1).

IEDB Mapping

In addition to these tools, the experimentally reported epitopes present in the Immune Epitope Database (IEDB) were also mapped on the vaccine target proteins selected for this study. **Figure 1** shows the complete workflow of this study.

Database Web Interface

All the predicted epitopes have been stored and presented on a web-portal named as VacTarBac. It is built on Apache HTTP server (version 2.2.17), which is installed on machine with Ubuntu as operating system. The responsive front-end, which is suitable for mobile, tablet, and desktop, was developed using HTML5, CSS3, PHP5, and JavaScript. MySQL (a relational database management system, version 5.0.51 b) was used at the back-end to manage the data.

RESULTS

Screening of Bacterial Species

The VFDB virulence factor proteins download link is available at http://www.mgc.ac.cn/VFs/Down/VFDB_setA_pro.fas.gz, providing all the sequences in FASTA format with the headers providing the information for the protein, for example, the VFDB id, source organism, etc. From these headers, the source organism names were extracted. The DEG is available for download at <http://www.essentialgene.org> in the form of FASTA sequences too. Only those sequences from DEG were extracted that sourced from organism species (indicated by the headers) present in the VFDB headers too. Finally, the protein sequences from the VFDB and DEG belonged to 14 bacterial species.

Screening of Target Proteins

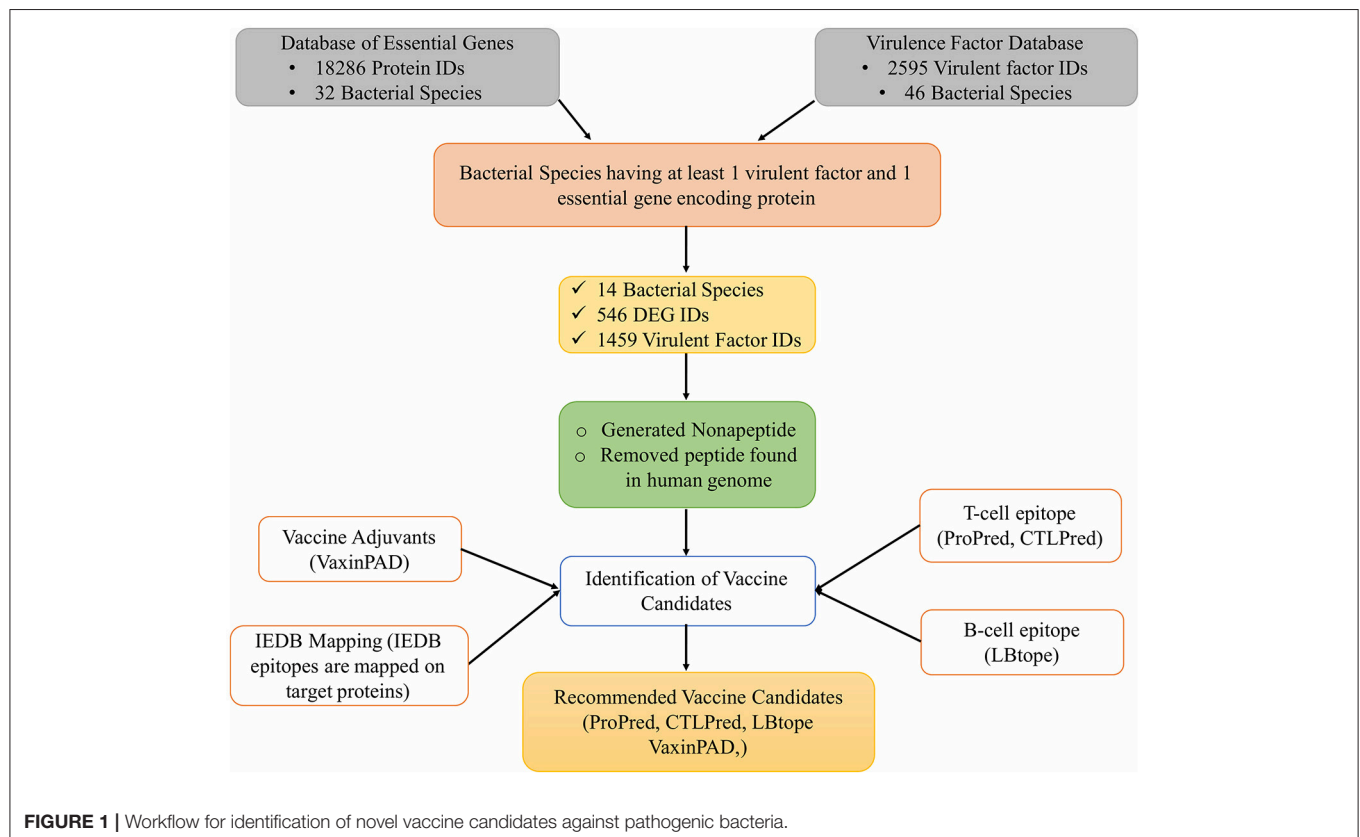
These 14 bacterial species contain 1459 virulence factor as stated in VFDB. All these 1459 virulent factors were selected

for the further study. Previous studies have advocated that bacterial proteins at specific cellular locations (20, 23–25, 33) or performing specific functions essential for cell (34) are potential candidates for designing subunit vaccine. Therefore, on the basis of function and localization within the cell, 546 Proteins from the 14 bacterial species, were categorized as membrane, envelope, repair, and secretory proteins. **Table 2** shows distribution of screened target proteins among selected 14 bacteria.

The most important aspect of this rational determination of vaccine targets was to extract proteins common in both VFDB and DEG dataset. Our analysis shows that 21 IDs from VFDB dataset were identical to 35 IDs of DEG dataset. (**Supplementary Table 1**) These are 21 proteins from 5 bacterial species. We consider these as the most suitable vaccine targets, being both, virulence factors as well as gene products essential for survival.

Epitope Predictions

After generating the nona-peptides, redundant nona-peptides as well as self-antigens were removed by mapping on 1000 human proteome. For rest of the peptides, we predicted the immunogenicity using our prediction pipeline. LBtope, ProPred, CTLpred, and VaxinPAD helped us identify B-cell epitopes, MHC class II binders, T-cell epitopes, and adjuvants respectively. These tools integrated and implemented as a pipeline helped identify the immunogenic regions of the individual proteins in the form of nona-mer epitopes. On the other hand, the visualization tools integrated in the platform “VacTarBac” helped envisage the



stretches of lengths longer than nona-mer having high density of predicted epitopes. The user may anticipate that these regions could prove to be antigens effective in the form of vaccines. Thus, for experimental investigation of the sequences that may prove to be effective vaccines, sequences longer than 9-mers containing numerous predicted epitopes could be taken up as a better strategy for the development of vaccines.

The epitopes predicted individually by each tool of our pipeline are too many to be tested experimentally. (Table 3) Therefore, we have applied an intuitive method to arrive at a reasonable number. (Figure 2 and Supplementary Table 2) Epitopes were separately identified for T-Cell and B-Cell categories. In case of T-Cell epitopes only those were retained that were predicted T-Cell epitopes, MHC-II Binders as well as adjuvants. For B-Cell epitopes, peptides predicted to be B-Cell epitopes, MHC-II Binders and adjuvants were finalized for recommendation as therapeutics.

Epitopes Assessed to be the Best Vaccine Candidates

The aim of this study is to enhance the vaccine designing against various pathogenic bacteria by predicting potential MHC binder, B cell and T cell epitopes as well as vaccine adjuvants using a prediction pipeline. Yet, the idea of using a pipeline instead of the prediction tools individually was intended to pose stepwise filters on the numerous peptides that may be predicted as epitopes for different arms of the immune system. These filters reduce the number of epitopes to be experimentally tested and lead to a final set of peptide sequences with more possibility of activating immune system as these are recommended by all the individual tools in the prediction pipeline. Epitopes positively predicted by all the prediction tools, provided coverage of 13 out of 14 bacterial species. An organism-wise as well as vaccine target protein category-wise representation as shown in Table 4 provides more comprehensive analysis of the epitopes assessed

TABLE 2 | Bacterial species-wise distribution of Target Proteins.

Bacteria	Virulent Factors	Protein encoded by Essential Genes				Virulent Factors which are Essential Gene products
		Membrane Protein	Repair Protein	Secretory Protein	Envelope Protein	
<i>Bacillus subtilis</i>	1	2	0	6	0	0
<i>Burkholderia pseudomallei</i>	148	27	0	4	0	0
<i>Campylobacter jejuni</i>	128	7	0	0	0	1
<i>Escherichia coli</i>	293	37	1	2	0	0
<i>Haemophilus influenzae</i>	85	10	4	0	0	0
<i>Helicobacter pylori</i>	114	8	0	0	12	1
<i>Mycobacterium tuberculosis</i>	65	241	3	23	0	14
<i>Pseudomonas aeruginosa</i>	242	18	10	8	0	2
<i>Salmonella enteric</i>	123	35	0	3	0	3
<i>Staphylococcus aureus</i>	80	11	13	6	12	0
<i>Streptococcus agalactiae</i>	45	1	0	0	0	0
<i>Streptococcus pneumoniae</i>	43	14	7	0	12	0
<i>Streptococcus pyogenes</i>	43	2	0	0	0	0
<i>Vibrio cholera</i>	49	7	0	0	0	0
Total	1459	420	38	52	36	21

TABLE 3 | Distribution of proteins, nona-peptides and predicted epitopes among various categories.

Source	VFDB		Database of Essential Genes (DEG)		
	Virulence factor	Membrane Protein	Repair Proteins	Secretory Protein	Envelope Protein
Number of Proteins	1459	420	38	52	36
Number of Nona-peptides	1058876	482420	37823	53517	30497
Number of unique Nona-peptides	244347	60436	9866	8935	6660
Number of exclusive Nona-peptides absent in 1000 human proteome	228589	60365	9860	8935	6645
T-Cell Epitopes	131159	33378	5225	5141	3504
B-Cell Epitopes	36758	7864	1494	1326	832
MHC-II Binders	87849	23968	3287	3207	2647
Adjuvants	47618	12259	2341	1882	1425

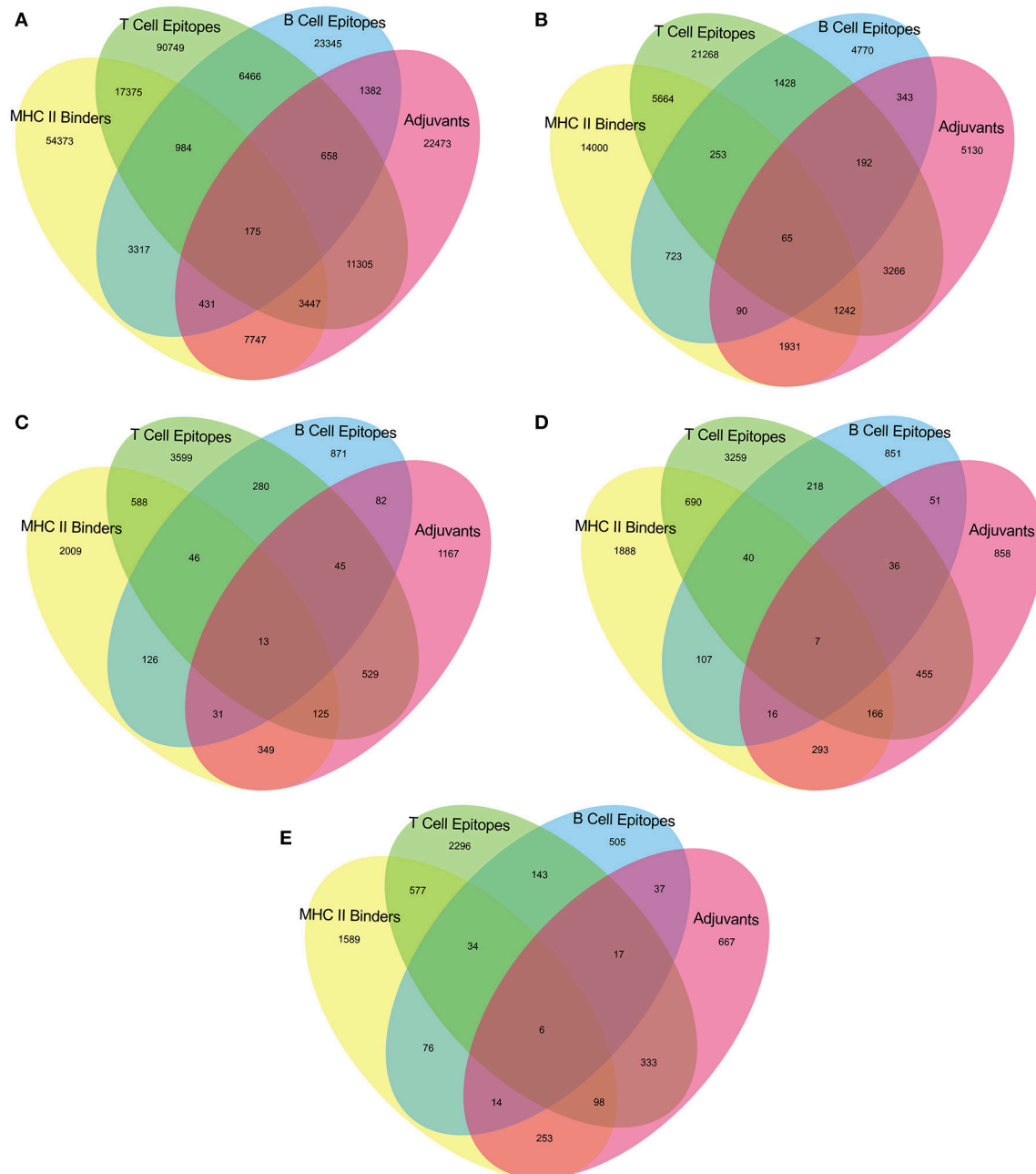


FIGURE 2 | Venn diagram representing number of predicted epitopes individually as well as by intuitive approach through prediction pipeline for (A) Virulence factors, and essential (B) Membrane proteins, (C) Repair proteins (D) Secretory proteins, and (E) Envelope proteins.

to be the “best vaccine candidate” using all the prediction tools.

Best Possible Antigenic Proteins

As stated earlier, the study started with 1459 virulence factor and 546 essential proteins, we have predicted epitopes from all of the generated nona-peptides. Some proteins, have high number of predicted MHC binders, T cell epitopes, B cell epitopes as well as vaccine adjuvants. For example, the peptide syntase (pyoverdine)

protein, a virulence factor of *Pseudomonas aeruginosa* results into 8197 predicted epitopes. Such proteins or regions within the proteins could be wisely used as potential antigens to stimulate the immune system (Supplementary Table 3). Figure 3 shows the portion of an essential protein from *Bacillus subtilis*, which has a high number of predicted epitopes. This indicates that instead of experimentally validating a single nona-peptide, taking a longer portion of the protein, having overlapping predicted epitopes, could prove to be a better approach.

TABLE 4 | Species-wise distribution of the best epitopes by filtration through all the *in silico* tools used in the study.

Bacteria	Best Epitopes (T-cell Epitope+B-Cell Epitope+MHC II Binder+Adjuvant)				
	Essential Gene Proteins				Virulent Factors
	Membrane	Repair	Secretory	Envelope	
<i>Bacillus subtilis</i>	0	0	0	0	0
<i>Burkholderia pseudomallei</i>	8	0	0	0	25
<i>Campylobacter jejuni</i>	0	0	0	0	17
<i>Escherichia coli</i>	7	0	1	0	42
<i>Haemophilus influenzae</i>	3	3	0	0	6
<i>Helicobacter pylori</i>	0	0	0	2	4
<i>Mycobacterium tuberculosis</i>	61	1	11	0	11
<i>Pseudomonas aeruginosa</i>	4	4	0	0	38
<i>Salmonella enterica</i>	2	0	0	0	15
<i>Staphylococcus aureus</i>	2	4	0	2	5
<i>Streptococcus agalactiae</i>	1	0	0	0	4
<i>Streptococcus pneumoniae</i>	3	1	0	2	3
<i>Streptococcus pyogenes</i>	1	0	0	0	3
<i>Vibrio cholerae</i>	0	0	0	0	6
Total number of Epitopes	92	13	12	6	179
Number of Unique Epitopes	65	13	7	6	175

Considering all the epitopes together in this table, these belonged to 13 bacterial species out of the total 14 species considered for the study.



FIGURE 3 | Mapping of predicted epitopes on one of the essential protein of *Bacillus subtilis* as **(A)** user friendly interactive Java-enabled view and **(B)** traditional simpler view. The blue colored sequences are the predicted 9-mer epitopes starting from red colored amino acid.

Webserver Implementation

To assist the scientific community in expediting the peptide-based vaccine designing against pathogenic bacteria, all the

predicted potential epitopes were compiled in the form of a database and were provided as a web-based service (<http://webs.iitd.edu.in/raghava/vactarbac/>). A user can browse all the

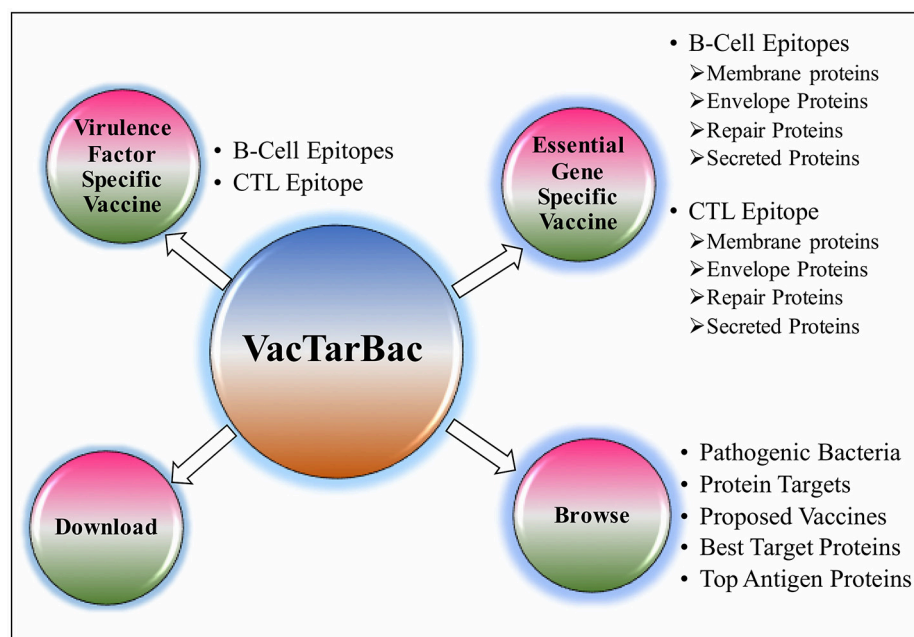


FIGURE 4 | Descriptive representation of VacTarBac webserver.

recommended potential B cell and T cell epitopes designed by targeting virulence factors as well as proteins encoded by essential genes of all 14 bacterial species. Browsing by the pathogenic bacteria as well as targeted proteins considered while designing of potential epitopes is also implemented. A user can also browse the top 5 antigens from the enlisted bacteria. In addition, a list of proposed vaccine candidates is also provided in the browse option (Figure 4). For the ease of user, results have been displayed in tabular and graphical form as well as an interactive visualization.

DISCUSSION AND CONCLUSION

The twentieth century witnessed a remarkable success in development of antibiotics against various pathogenic bacteria but the emergence of drug resistance and toxicity caused a shift in the antimicrobial strategies toward peptide-based therapies. Several anti-bacterial peptides have been studied to check the pestiferous effects of pathogenic bacteria (35, 36). Rapid bacteriocidal activity and low propensity for resistance development are some of the major assets of these peptides, but high cost, limited stability and unknown toxicology, and pharmacokinetics are the major disadvantages (37). Contrary to antibiotics or peptide-based drugs, instigating host immune responses against the pathogen offers multifaceted management of the invasion and has been successfully achieved in the past by the use of vaccines like DPT, BCG etc against few deadly infectious diseases like tetanus, diphtheria, tuberculosis etc. Consequently, active search for vaccines is underway currently for many infectious diseases caused by pathogenic bacteria, for example, melioidosis caused by *Burkholderia pseudomallei* (2).

Even in case of tuberculosis, for which multiple lines of drugs and BCG vaccines are available; reduced efficacy of BCG and antibiotic resistance forced investigation of better and more potent vaccines. A recent computational effort has gone up to the strain level comparison and considered proteins as vaccine targets that were shared among tuberculoid, non-tuberculoid, and vaccine strains (16). Continuous efforts are being made by researchers to identify novel vaccine candidates against several pathogens such as *Ebola* virus (14), *Zika* virus (15), *vaccinia* virus (38), *Neisseria meningitidis* (39), *Corynebacterium pseudotuberculosis* (40) *Edwardsiella tarda*, and *Flavobacterium columnare*; fish pathogens (41, 42).

In past, several *in silico*, protein-based vaccine candidates have been identified, using peptide conservation score and predicted peptide-MHC binding (43) as well as molecular docking and MHC-peptide complex stabilization assay (18). In this study, we employed immunoinformatic tools to identify antigen and epitope-based vaccine candidates, having the potential to evoke one of the many arms of immune system. The present study includes proteins; virulence factors as well as proteins encoded by genes essential for survival from multiple pathogenic bacteria. The pipeline created for the identification of vaccine candidates, consists of widely used, accurate and recommended tools (12, 13). Our pipeline also includes a tool to identify peptide-based adjuvant candidates, a major advantage over other studies. The numbers of epitopes predicted by the individual tools within the pipeline that are not present in the proteome derived from the 1000 human genomes, were in thousands. For arriving at economical number of epitopes that could be tested experimentally, an intuitive rational criterion was to extract epitopes predicted positive by more than one tool in the pipeline. This would effectively mean that the selected epitope could have

a predictive capability of evoking different arms of the immune system.

Upon executing the rational criteria, 1459 virulent factors from VFDB yielded 3622 nona-peptides having ability to bind with MHC-II, activate T-cell response and could act as self-adjuvants. Beside this, peptides are predicted to bind with MHC-II, activate B-cell response as well as self-adjuvants. Similarly, 420 membrane proteins from the DEG provided 1307 predicted T-cell epitopes and 155 predicted B-cell epitopes that are also MHC II binders and adjuvants. The 36 DEG envelope proteins when subjected to the epitope prediction pipeline, yield 104 predicted T-cell epitopes and 20 predicted B-cell epitopes with probability that these will bind MHC Class II molecules and would also be adjuvants. In cases of 52 secretory proteins and 38 repair proteins taken from the DEG, the predicted T-cell epitopes, and B-cell epitopes were respectively 173 and 23 for secretory proteins and 138 and 44 for the repair proteins, all of them being also the positively predicted MHC II binders as well as adjuvants.

In conclusion, the ultimate aim of the present study was the identification of epitopes capable of activating multiple wings of the human immune system arrived at by filtering the epitopes through stringent criteria. Combining the results of the VFDB and DEG datasets, this study was able to identify 252 unique epitopes predicted to be T-cell epitopes, B-cell epitopes, MHC II binders and Vaccine Adjuvants not present in the human proteome (belonging to the 1000 genomes) extracted from proteins (essential gene products and/or virulence factors) of 13 bacterial species. All the recommended vaccine candidates have been stored in a repository; VacTarBac (<http://webs.iitd.edu.in/raghava/vactarbac/>) freely available on the world-wide web.

Peptides are the promising candidate as immunotherapeutic but are less immunogenic when used alone as vaccine. They need potent immunostimulatory adjuvants to effectively activate the innate and adaptive arms of immune system. In this study, we have implemented VaxinPAD in our prediction pipeline, which aids in predicting based vaccine adjuvant, thus strengthening the study. Beside this, peptide formulation is a critical task and a formulation scientist must overcome the chemical instability of peptides. The conformational fluxionality and propensity to self-associate makes peptide more difficult to formulate. Various computational tools predict the aggregation propensities of polypeptide chains such as Zyggregator (44), PASTA 2.0 (45), etc. These can be used prior to formulating the peptide-based vaccine. THPdb is a database of FDA approved protein and peptide therapeutics, and provide detailed formulation strategies of peptide-based drug available in the market (46). Although little literature exist about formulating peptide drug product but strategies used in successful peptide drugs will be helpful in designing newer formulation techniques and constituent for peptide based vaccines.

REFERENCES

1. Neill JO'. *Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations The Review on Antimicrobial Resistance Chaired* (2014).

We have predicted epitopes based on highly cited, published and accurate immune epitope prediction tools. Yet, these prediction algorithms have their own limitations. Thus, the antigen or epitope should be experimentally validated before suggesting it for medical purpose. Apart from this, the platform, VacTarBac, would require continuous revamp owing to the rapidly growing sequencing data particularly that of the novel and emerging pathogenic strains. A future work for this platform may include comparison of the pathogenic proteomes at the strain level. This may lead to identification of recently acquired proteins/peptides of the pathogen that may have rendered existing therapeutic strategies against the emergent strain ineffective. Moreover, the platform, VacTarBac, still lacks information of vaccine formulations that may be added in future to help in the actual development of the vaccines. Despite such limitations, we anticipate that the current study and the current data in the repository VacTarBac, will be helpful for researchers and will boost and hasten the vaccine designing against pathogenic bacteria considered in the study.

AVAILABILITY AND UPDATE OF THE RESOURCE

VacTarBac is freely available at <http://webs.iitd.edu.in/raghava/vactarbac/>. We will update the platform at every 6 months depending upon the availability of virulence factor and essential genes information in other resources.

AUTHOR CONTRIBUTIONS

SU and GN downloaded and processed the data. GN prepared the pipeline and predicted epitopes. SU and GN analyzed results and prepared tables and figures. SU, GN, and GR wrote the manuscript. SU developed web interface. GR conceived the idea and coordinated the project.

FUNDING

Authors are thankful to funding agencies J.C. Bose National Fellowship (DST), Department of Biotechnology (DBT) and Council of Scientific and Industrial research (CSIR) for fellowships and financial support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2018.02280/full#supplementary-material>

2. Peacock SJ, Limmathurotsakul D, Lubell Y, Koh GCKW, White LJ, Day NPJ, et al. Melioidosis vaccines: a systematic review and appraisal of the potential to exploit biodefense vaccines for public health purposes. *PLoS Negl Trop Dis.* (2012) 6:e1488. doi: 10.1371/journal.pntd.0001488

3. Andersen P, Doherty TM. The success and failure of BCG - implications for a novel tuberculosis vaccine. *Nat Rev Microbiol.* (2005) 3:656–62. doi: 10.1038/nrmicro1211
4. Kernodle DS. Decrease in the effectiveness of Bacille Calmette-Guérin vaccine against pulmonary tuberculosis: a consequence of increased immune suppression by microbial antioxidants, not overattenuation. *Clin Infect Dis.* (2010) 51:177–84. doi: 10.1086/653533
5. Barnett TC, Lim JY, Soderholm AT, Rivera-Hernandez T, West NP, Walker MJ. Host-pathogen interaction during bacterial vaccination. *Curr Opin Immunol.* (2015) 36:1–7. doi: 10.1016/j.coi.2015.04.002
6. Zuñiga S, Pascual-Iglesias A, Sanchez CM, Sola I, Enjuanes L. Virulence factors in porcine coronaviruses and vaccine design. *Virus Res.* (2016) 226:142–51. doi: 10.1016/j.virusres.2016.07.003
7. Sánchez J, Holmgren J. Virulence factors, pathogenesis and vaccine protection in cholera and ETEC diarrhea. *Curr Opin Immunol.* (2005) 17:388–98. doi: 10.1016/j.coi.2005.06.007
8. Chen L, Xiong Z, Sun L, Yang J, Jin Q. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Res.* (2012) 40:D641–5. doi: 10.1093/nar/gkr989
9. Lu Y, Deng J, Rhodes JC, Lu H, Lu LJ. Predicting essential genes for identifying potential drug targets in *Aspergillus fumigatus*. *Comput Biol Chem.* (2014) 50:29–40. doi: 10.1016/j.compbiolchem.2014.01.011
10. Hu W, Sillaots S, Lemieux S, Davison J, Kauffman S, Breton A, et al. Essential gene identification and drug target prioritization in *Aspergillus fumigatus*. *PLoS Pathog.* (2007) 3:e24. doi: 10.1371/journal.ppat.0030024
11. Luo H, Lin Y, Gao F, Zhang CT, Zhang R. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res.* (2014) 42:D574–80. doi: 10.1093/nar/gkt1131
12. Usmani SS, Kumar R, Bhalla S, Kumar V, Raghava GPS. *In silico* tools and databases for designing peptide-based vaccine and drugs. *Adv Protein Chem Struct Biol.* (2018) 112:221–63. doi: 10.1016/bs.apcsb.2018.01.006
13. Dhanda SK, Usmani SS, Agrawal P, Nagpal G, Gautam A, Raghava GPS. Novel *in silico* tools for designing peptide-based subunit vaccines and immunotherapeutics. *Brief Bioinformatics* (2017) 18:467–78. doi: 10.1093/bib/bbw025
14. Dhanda SK, Chaudhary K, Gupta S, Brahmachari SK, Raghava GPS. A web-based resource for designing therapeutics against Ebola Virus. *Sci Rep.* (2016) 6:24782. doi: 10.1038/srep24782
15. Gupta AK, Kaur K, Rajput A, Dhanda SK, Sehgal M, Khan MS, et al. ZikaVR: an integrated zika virus resource for genomics, proteomics, phylogenetic and therapeutic analysis. *Sci Rep.* (2016) 6:32713. doi: 10.1038/srep32713
16. Dhanda SK, Vir P, Singla D, Gupta S, Kumar S, Raghava GPS. A web-based platform for designing vaccines against existing and emerging strains of *Mycobacterium tuberculosis*. *PLoS ONE* (2016) 11:e0153771. doi: 10.1371/journal.pone.0153771
17. Danielson RE, Cooper RC. The effect of water activity on *Legionella* spp. growth and survival. *J Appl Bacteriol.* (1986) 61:157–61.
18. Zheng J, Lin X, Wang X, Zheng L, Lan S, Jin S, et al. *In silico* analysis of epitope-based vaccine candidates against hepatitis B virus polymerase protein. *Viruses* (2017) 9:112. doi: 10.3390/v9050112
19. Meza B, Ascencio F, Sierra-Beltrán AP, Torres J, Angulo C. A novel design of a multi-antigenic, multistage and multi-epitope vaccine against *Helicobacter pylori*: an *in silico* approach. *Infect Genet Evol.* (2017) 49:309–17. doi: 10.1016/j.meegid.2017.02.007
20. Rauta PR, Ashe S, Nayak D, Nayak B. *In silico* identification of outer membrane protein (Omp) and subunit vaccine design against pathogenic *Vibrio cholerae*. *Comput Biol Chem.* (2016) 65:61–8. doi: 10.1016/j.compbiolchem.2016.10.004
21. Fahimi H, Sadeghizadeh M, Mohammadipour M. *In silico* analysis of an envelope domain III-based multivalent fusion protein as a potential dengue vaccine candidate. *Clin Exp Vaccine Res.* (2016) 5:41–9. doi: 10.7774/cevr.2016.5.1.41
22. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* (2016) 44:D694–7. doi: 10.1093/nar/gkv1239
23. Hajighahramani N, Nezafat N, Eslami M, Negahdaripour M, Rahmatbadi SS, Ghasemi Y. Immunoinformatics analysis and *in silico* designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*. *Infect Genet Evol.* (2017) 48:83–94. doi: 10.1016/j.meegid.2016.12.010
24. Yasmin T, Akter S, Debnath M, Ebihara A, Nakagawa T, Nabi AHMN. *In silico* proposition to predict cluster of B- and T-cell epitopes for the usefulness of vaccine design from invasive, virulent and membrane associated proteins of *C. jejuni*. *In silico Pharmacol.* (2016) 4:5. doi: 10.1186/s40203-016-0020-y
25. Alam A, Ali S, Ahamad S, Malik MZ, Ishrat R. From ZikV genome to vaccine: *in silico* approach for the epitope-based peptide vaccine against Zika virus envelope glycoprotein. *Immunology* (2016) 149:386–99. doi: 10.1111/imm.12656
26. Li M, Song S, Yang D, Li C, Li G. Identification of secreted proteins as novel antigenic vaccine candidates of *Haemophilus parasuis* serovar 5. *Vaccine* (2015) 33:1695–701. doi: 10.1016/j.vaccine.2015.02.023
27. Blander SJ, Horwitz MA. Vaccination with the major secretory protein of *Legionella* induces humoral and cell-mediated immune responses and protective immunity across different serogroups of *Legionella pneumophila* and different species of *Legionella*. *J Immunol.* (1991) 147:285–91.
28. Green ER, Mecsas J. Bacterial secretion systems: an overview. *Microbiol Spectr.* (2016) 4:1236–7. doi: 10.1128/microbiolspec.VMBF-0012-2015
29. Žgur-Bertok D. DNA damage repair and bacterial pathogens. *PLoS Pathog.* (2013) 9:e1003711. doi: 10.1371/journal.ppat.1003711
30. Singh H, Ansari HR, Raghava GPS. Improved method for linear B-cell epitope prediction using antigen's primary sequence. *PLoS ONE* (2013) 8:e62216. doi: 10.1371/journal.pone.0062216
31. Singh H, Raghava GP. ProPred: prediction of HLA-DR binding sites. *Bioinformatics* (2001) 17:1236–7. doi: 10.1093/bioinformatics/17.12.1236
32. Bhasin M, Raghava GPS. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine* (2004) 22:3195–204. doi: 10.1016/j.vaccine.2004.02.005
33. Rana A, Thakur S, Bhardwaj N, Kumar D, Akhter Y. Excavating the surface-associated and secretory proteome of *Mycobacterium leprae* for identifying vaccines and diagnostic markers relevant immunodominant epitopes. *Pathog Dis.* (2016) 74:ftw110. doi: 10.1093/femspd/ftw110
34. Damte D, Suh JW, Lee SJ, Yohannes SB, Hossain MA, Park SC. Putative drug and vaccine target protein identification using comparative genomic analysis of KEGG annotated metabolic pathways of *Mycoplasma hyopneumoniae*. *Genomics* (2013) 102:47–56. doi: 10.1016/j.ygeno.2013.04.011
35. Usmani SS, Kumar R, Kumar V, Singh S, Raghava GPS. AntiTbPdb: a knowledgebase of anti-tubercular peptides. *Database* (2018) 2018. doi: 10.1093/database/bay025
36. Wang G, Li X, Wang Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* (2016) 44:D1087–93. doi: 10.1093/nar/gkv1278
37. Gordon YJ, Romanowski EG, McDermott AM. A review of antimicrobial peptides and their therapeutic potential as anti-infective drugs. *Curr Eye Res.* (2005) 30:505–15. doi: 10.1080/02713680590968637
38. Sánchez-Sampedro L, Gómez CE, Mejías-Pérez E, Pérez-Jiménez E, Oliveros JC, Esteban M. Attenuated and replication-competent vaccinia virus strains M65 and M101 with distinct biology and immunogenicity as potential vaccine candidates against pathogens. *J Virol.* (2013) 87:6955–74. doi: 10.1128/JVI.03013-12
39. Christodoulides M, Heckels J. Novel approaches to *Neisseria meningitidis* vaccine design. *Pathog Dis* (2017) 75:ftx033. doi: 10.1093/femspd/ftx033
40. Barh D, Jain N, Tiwari S, Parida BP, D'Afonseca V, Li L, et al. A novel comparative genomics analysis for common drug and vaccine targets in *Corynebacterium pseudotuberculosis* and other CMN group of human pathogens. *Chem Biol Drug Des* (2011) 78:73–84. doi: 10.1111/j.1747-0285.2011.01118.x
41. Liu X, Sun J, Wu H. Glycolysis-related proteins are broad spectrum vaccine candidates against aquacultural pathogens. *Vaccine* (2017) 35:3813–6. doi: 10.1016/j.vaccine.2017.05.066
42. Mahendran R, Jeyabaskar S, Sitharaman G, Michael RD, Paul AV. Computer-aided vaccine designing approach against fish pathogens *Edwardsiella tarda* and *Flavobacterium columnare* using bioinformatics softwares. *Drug Des Devel Ther.* (2016) 10:1703–14. doi: 10.2147/DDDT.S95691
43. Goodswen SJ, Kennedy PJ, Ellis JT. Enhancing *in silico* protein-based vaccine discovery for eukaryotic pathogens using predicted peptide-MHC

- binding and peptide conservation scores. *PLoS ONE* (2014) 9:e115745. doi: 10.1371/journal.pone.0115745
44. Tartaglia GG, Vendruscolo M. The Zyggregator method for predicting protein aggregation propensities. *Chem Soc Rev.* (2008) 37:1395–401. doi: 10.1039/b706784b
45. Walsh I, Seno F, Tosatto SC, Trovato A. PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Res.* (2014) 42:W301–7. doi: 10.1093/nar/gku399
46. Usmani SS, Bedi G, Samuel JS, Singh S, Kalra S, Kumar P, et al. THPdb: Database of FDA-approved peptide and protein therapeutics. *PLoS ONE* (2017) 12:e0181748. doi: 10.1371/journal.pone.0181748

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Nagpal, Usmani and Raghava. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Bacterial Vaccine Antigen Discovery in the Reverse Vaccinology 2.0 Era: Progress and Challenges

Fadil A. Bidmos*, Sara Siris, Camilla A. Gladstone and Paul R. Langford

Department of Medicine, Imperial College London, London, United Kingdom

OPEN ACCESS

Edited by:

Pedro A. Reche,
Complutense University of Madrid,
Spain

Reviewed by:

Sudheer Gupta,
All India Institute of Medical Sciences
Bhopal, India
Pietro Speziale,
Università degli Studi di Pavia, Italy
Paola Massari,
Tufts University School of Medicine,
United States

*Correspondence:

Fadil A. Bidmos
f.bidmos@imperial.ac.uk

Specialty section:

This article was submitted to
Vaccines and Molecular Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 30 June 2018

Accepted: 17 September 2018

Published: 08 October 2018

Citation:

Bidmos FA, Siris S, Gladstone CA and
Langford PR (2018) Bacterial Vaccine
Antigen Discovery in the Reverse
Vaccinology 2.0 Era: Progress and
Challenges. Front. Immunol. 9:2315.
doi: 10.3389/fimmu.2018.02315

The ongoing, and very serious, threat from antimicrobial resistance necessitates the development and use of preventative measures, predominantly vaccination. Polysaccharide-based vaccines have provided a degree of success in limiting morbidity from disseminated bacterial infections, including those caused by the major human obligate pathogens, *Neisseria meningitidis*, and *Streptococcus pneumoniae*. Limitations of these polysaccharide vaccines, such as partial coverage and induced escape leading to persistence of disease, provide a compelling argument for the development of protein vaccines. In this review, we briefly chronicle approaches that have yielded licensed vaccines before highlighting reverse vaccinology 2.0 and its potential application in the discovery of novel bacterial protein vaccine candidates. Technical challenges and research gaps are also discussed.

Keywords: reverse vaccinology 2.0, human monoclonal antibodies, bacterial pathogens, vaccine candidate antigens, immunotherapy

BACKGROUND: VACCINE DISCOVERY IN THE PRE-WHOLE GENOME SEQUENCING (WGS) ERA

Precedent to advancements in genomics, vaccines were developed based on Pasteur's *rules of vaccinology*. A 230-fold serial passage of a bovine bacillus in bile medium produced the live attenuated Bacillus Calmette-Guerin (BCG) vaccine against *Mycobacterium tuberculosis* (MTB) (1). Then, a trivalent blend of poliovirus inactivated in <0.5% formalin was used by Salk et al. as a safe and effective vaccine against poliovirus (2). The inability to culture some pathogens *in vitro* (owing to safety or lack of suitable culture conditions), extensive antigenic variability, and molecular mimicry limit the broad applicability of traditional culture-based techniques in the development of vaccines targeting other economically-important pathogens such as *Mycobacterium leprae* and *Neisseria meningitidis*.

The inadequacies of culture-based techniques caused a shift in focus to the use of subunit components as vaccine candidates. Identification of these subunit vaccine candidates was largely hypothesis-driven, targeted cellular components and were often well-known virulence factors: for example, the pertussis toxin and fimbriae in the acellular pertussis vaccine (3); and the meningococcal porin, PorA, in the epidemic-specific detergent-extracted outer membrane vesicle (OMV) vaccines of Chile, Norway, and New Zealand (4–6). With complex biosynthetic methods, bacterial capsular polysaccharides served as prime components of effective vaccines, used singly, for example, in the 23-valent pneumococcal polysaccharide vaccine, PPSV23 (7). In addition, these capsular antigens have been conjugated to carrier proteins in a dose-sensitive manner for enhanced immunogenicity, as found in the *Haemophilus influenzae* type b (Hib) (8), pneumococcal

(9), and meningococcal (10) conjugate vaccines. The limited capacity of these hypothesis-driven studies to focus on only a handful of candidates at a time was costly in time, labor and financial terms. This is especially because of the large pool from which prospective candidates for individual bacterial pathogens are screened, coupled with the low likelihood of targets satisfying key vaccine candidacy criteria (abundantly-expressed, surface-exposed, functionally-immunogenic, and highly-conserved). Thus, alternative high throughput methods were sought to accelerate the pre-clinical vaccine development phase, especially in situations requiring rapid curtailment of disease transmission.

WHOLE GENOMIC AND PROTEOMIC APPROACHES

Reverse Vaccinology (RV)

The publication of the first complete bacterial genome sequence in 1995 [for *H. influenzae* (11)] heralded a revolution in approaches to vaccine development. By using genomic data and preset bioinformatic screens, putative surface-associated antigens of a pathogen were identified. The subsequent recombinant expression of these genes and immunization of animals with recombinant proteins, for the determination of active and passive levels of protection, provided data that substantiated or annulled the vaccine candidacy of selected antigens (12, 13). This “classical” RV approach led to the development of the multicomponent meningococcal serogroup B vaccine (4CMenB) (14). While 4CMenB has potential for cross-serogroup protection (15), it has been argued that pan-genomic *in silico* analysis is more appropriate because of the high degree of intraspecific diversity exhibited by many bacterial pathogens (16). Using this pan-genomic approach, Maione et al. (17) identified four protective antigens from the analysis of an octa-genomic panel derived from the most prevalent disease-causing *Streptococcus agalactiae* strains. The main attraction of RV lies in its applicability to any pathogen with WGS data and to which antibody-mediated immunity for protection against disease is crucial. Its use in the discovery of candidate antigens comprising vaccines targeting other bacterial pathogens, including the multidrug-resistant *Acinetobacter baumannii*, has been demonstrated (18–20). However, important non-classical surface-associated proteins may be missed due to the parameters of the bioinformatic screen(s).

Related to RV is the use of transcriptomics to identify novel vaccine antigens. For example, the comparative analysis of the meningococcal transcriptome in *ex vivo* human whole blood and *in vitro* nasopharyngeal colonization models revealed three antigens that were differentially regulated between invasive disease and asymptomatic colonization, and were thus subjects for further vaccine candidacy studies (21). However, this transcriptomics-based approach has not been widely employed.

Surfome and Secretome Analysis

Whole proteomic approaches, involving enzymatic processing of whole cells or extracellular exudates followed by liquid-chromatography mass spectrometry (LC-MS) or peptide fragment fingerprinting, also allow for high-throughput

screening of the antigenic repertoire of a pathogen (22). The power of these proteomic methods in identifying rare protective antigens missed by the *in silico* screens of RV makes them appealing [as exemplified by the case of the cell wall-anchored antigen, SAN_1485, of *S. agalactiae* (23)]. Converse to RV, proteolytic digestion is more suited toward Gram-positive bacteria, since Gram-negative bacteria are more susceptible to proteolysis-induced cell lysis.

REVERSE VACCINOLOGY 2.0

The majority of currently-available bacterial vaccines provide protection by inducing pathogen-specific antibodies. Therefore, harnessing the antibody component of a potent human humoral response to disseminated infection is valuable for the identification of novel protective antigens. This approach, termed reverse vaccinology 2.0 (RV 2.0) (24, 25), relies on the isolation and recombinant expression of the variable regions of heavy (VH) and light (VL = κ or λ) chain genes of immunoglobulin (focus has centered on IgG) using a variety of molecular tools. Enriched by the development of high-throughput technologies, the screening of large numbers of antibody-secreting cells (ASCs) is also advancing knowledge of host-pathogen interactive biology and auto-immunity (26, 27).

Monoclonal Antibody (mAb) Generation From ASCs

The first, and perhaps most crucial, phase of RV 2.0 is the cloning of human monoclonal antibodies (mAbs) from ASCs. Previously, immortalization of these ASCs via myeloma fusions or Epstein Barr virus (EBV) transformation were valuable to mAb production (28, 29). Because these were culture-based methods, the survival of all B-cells was not guaranteed and the omission of ASCs expressing antibodies cognate to crucial antigens was probable. Other techniques such as phage-display technology (30) and proteomic mining (31, 32) circumvent the unique issues affecting ASC immortalization techniques by focusing on recombinant antibody expression. However, the small proportion of antigen-specific antibodies (estimated at 10–15%) that are produced (33) because of the random pairing of VH and VL sequences make phage display and proteomic mining imprecise.

A more favored approach to mAb cloning is the single-cell sorting of ASCs into multi-well plates using flow cytometry, followed by the cloning of mAbs from each well (34, 35). To clone a high proportion of antigen-specific antibodies, this approach, termed expression cloning, requires blood sampling during the peak immune response and is thus more suited to short-lived plasmablasts (CD3⁺, CD14⁺, CD19⁺, CD20⁺, CD56⁺, CD27^{high}, and CD38^{high}), since higher circulating numbers of these are indicative of very recent history of infection (36). Notwithstanding, several studies have demonstrated its applicability to memory B-cells (37, 38). Further *in vitro* selection of antigen-specific plasmablasts or memory B-cells using eGFP-bound viral-like particles (39), labeled-antigen probes (40, 41) or *in vivo* antigen-specific plasmablast enrichment

in irradiated SCID/beige chimera mice (42) enhance the pathogen-specific mAb output of the approach. Converse to phage display and proteomic mining technologies, expression cloning yields mAbs with natural, host-like VH+VL pairings. Further refinements to this elegant method include: substituting restriction endonuclease cloning with Gibson assembly to enhance cloning precision (43); assembly of both VH and VL fragments into a single expression vector (44); and succeeding cell sorting with paired-chain antibody repertoire sequencing, thereby encompassing all V gene families, including unique clones expressed at low frequencies (33, 45, 46).

Assessment of Recombinant mAb Function

Subsequent to cloning, the clinical relevance of mAbs is assessed in *in vivo* investigations of passive immunity (47) or *in vitro* functional assays: for example, the well-established viral neutralization (48) and serum bactericidal assays (49), some of which have provided data employed in vaccine licensure (50). The cognate antigens targeted by functional mAbs can subsequently be determined using protein array screens or classical immunoproteomic approaches.

Application of RV 2.0 to Viral Vaccine Development

The power of RV 2.0 (see **Figure 1**) in the identification of viral vaccine candidates has been demonstrated in several studies focussing on human cytomegalovirus (HCMV), respiratory syncytial virus (RSV), HIV, influenza and dengue viruses (25, 51). Some of these candidate antigens, discovered using RV 2.0, include a novel pentameric glycoprotein complex, the gHgLPUL128L pentamer, which induces high neutralizing titres against HCMV in mice (52) and the F protein of RSV stabilized to the prefusion conformation (53, 54). Accruing data from phase 1b/2a clinical trials show that a mAb (MEDI8897) reactive with prefusion F epitopes is effective when used prophylactically in preterm infants (55). Like MEDI8897, mAb MHAA4549A, cloned from a healthy vaccinee and which targets and neutralizes all known influenza A strains (56), demonstrated significant antiviral activity in a phase 2 human influenza A virus challenge (57). Thus, these studies have signified the use of RV 2.0 in producing broadly-neutralizing mAbs for post-infection prophylaxis in addition to identifying functionally-immunogenic vaccine candidates.

POTENTIAL APPLICATION TO ANTIBACTERIAL VACCINE DISCOVERY

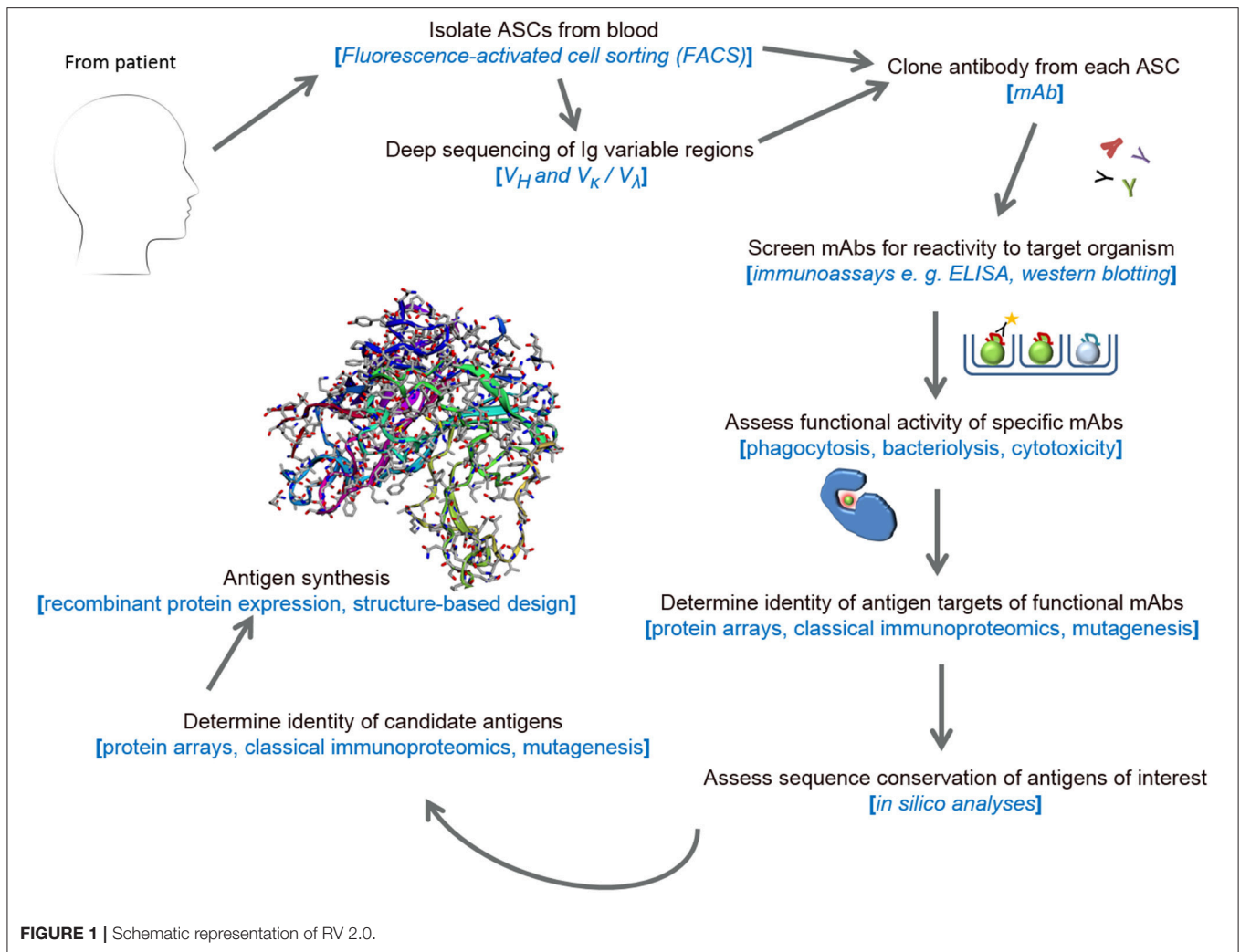
Judging by the progress made with the development of novel and effective viral immunotherapies, RV 2.0 is showing promise and is equally applicable to bacterial vaccinology. RV 2.0 was employed by Lu et al. (58) to identify functional anti-*Staphylococcus aureus* mAbs induced during bacteraemia. A total of ten mAbs were produced, four of which enhanced opsonophagocytosis of Wood46, a *S. aureus* reference strain. While three of the four functional mAbs targeted *S. aureus* antigens with known identities, the fourth mAb reacted

with a novel antigen. Recently, Zimmermann et al. (59) also demonstrated that functional anti-MTB surface antigen antibodies can be cloned from patient-derived plasmablasts of reactivated memory B-cell origins, providing further evidence for a role for antibodies in the modulation of potent immune responses toward MTB. Taken together with other studies investigating the importance of antibody-mediated neutralization of intracellular pathogens (60), a role for the vaccine-induced generation of antibodies against pathogens such as MTB and *Chlamydia trachomatis*, using antigens derived with RV 2.0 is, thus, evidenced. Similarly, Bidmos et al. (61) and Blum et al. (45) cloned functional antibodies from sufferers of meningococcal and Lyme disease, respectively; thus, underscoring the utility of the approach for identifying novel targets in different classes of bacteria. Continued use of RV 2.0 in bacterial vaccine discovery is, therefore, encouraged following the surmounting of technical challenges and filling of research gaps. In the following sections of this mini-review, emphasis will be placed on human mAb cloning and serological correlates of protection, since other related technical aspects of RV 2.0 such as recombinant protein expression, high-throughput sequencing of bacterial genomes and antibody repertoires, antigen identity determination and structure-based antigen design have been reviewed elsewhere (62–66).

Pathogen-Specific mAb Output

To identify novel antigens using the expression cloning method with precision, plasmablasts from patients convalescing from bacterial disease are required. Fundamental to the application of the expression cloning approach, therefore, is the determination of the magnitude and peak duration of the plasmablast response in these patients. The information on the duration of peak plasmablast circulation instructs optimum sampling time, which in turn impacts on the precision of pathogen-specific mAb generation. Studies assessing this magnitude of circulating plasmablast following bacterial infection have reported similar durations of peak response to those reported for primary or secondary viral infections [6–7 days for primary infections and ~10 days post-infection for secondary infections; reviewed in (36)]. Recently, Band et al. (67) reported a significant induction of differentiating (Ki-67⁺) plasmablasts in patients of nosocomial bacterial infections compared to healthy controls. This induction peaked between days 8 and 16 post-culture positivity in *A. baumannii*-infected patients reaching levels as high as 21% of the total lymphocyte population. Perhaps unsurprisingly, it was also observed that this induction was markedly different in individuals, reflecting differences in immunocompetence, as peak plasmablast levels ranged from: 1 to 21% among *A. baumannii*-infected patients; and 5–40% in *Escherichia coli*-infected patients. Consistent with the findings of Band et al. (67), a plasmablast response presented to: MTB infection in 38% of a patient cohort with levels ranging from 1 to 4% in those with strong serum IgG responses (59); *S. aureus* bacteraemia with mean levels of ~3.2% (1–7% range) (58); and up to 4% of circulating CD19⁺ cells in untreated sufferers of Lyme borreliosis (45).

The implications of these data for the precision of the pathogen-specific mAb output are considerable. Firstly, there



is a paucity of information in published literature on how many plasmablasts are pathogen-specific (bacterial) following inductions in patients, owing to the unavailability of suitable molecular probes that will enhance the Fluorescent Activated Cell Sorting (FACS) gating strategy. In the absence of such data, strategies such as the Ig-capture based technique described by Pinder et al. (40) could be employed to enrich for specific plasmablasts. It is more likely that complex antigens (whole bacterial cells, OMV, or outer membrane preparations) would be more beneficial in these strategies, when adapted, compared to single-antigen probes in order to obtain a plasmablast population targeting a wider antigen pool. It is noteworthy that in cases where patients are subjected to immediate antibiotic therapy on hospital admission because of rapid progression of disease (e.g., septicemia and meningitis), clinical isolates may be unobtainable, making the design of plasmablast enrichment probes difficult (also, the reason behind the unsuitability of memory B-cells in the absence of enrichment strategies). While clinical isolates from other disease sufferers could be utilized, they are non-ideal because rare mAb epitopes specific for the

infecting strain will be missed. Secondly, considering differences in the magnitude of the plasmablast response and for logistic reasons (for example, restrictions on blood sample volume in pediatric cases), pooling of patient samples may be required for the generation of a highly-diverse plasmablast pool, targeting several antigens, (and their variants) of the same pathogen. This is especially necessary for pathogens in which certain antigens are immunodominant such as PorA of *N. meningitidis*, which may mask immunity to rare but equally protective antigens.

If a total plasmablast sort approach is warranted (i.e., inclusive of non-pathogen specific plasmablasts), an attractive option is the rational selection of over-represented V_H+V_L combinations for mAb cloning based on the assumption that overrepresentation of V families, specifically among plasmablasts, is an indicator of preferential usage in response to a pathogen. Adequate depth of sequencing is, however, required in order to avoid non-inclusion of clonal V families expressed at lower frequencies (58). *In silico* analysis should also include antibodies with similar complementarity-determining region H3 loops [key to antibody conformation and affinity (68)] in addition to the exploitation

of de-noising algorithms, which would minimize the presence of errors introduced by sequencing (69).

In vitro Serological Correlates of Protection

Assessment of pathogen-specific mAb function is performed via standardized assays. Given the differences in biology of bacterial pathogens, these assays are specifically tailored to reflect mode of clearance of the pathogen from systemic circulation. Antibody-driven, complement-dependent bactericidal activity is measured in the standardized serum bactericidal assays designed for the meningococcus (49) while phagocytosis by neutrophils and macrophages enhanced by opsonic antibodies is assessed in the opsonophagocytic assays used in pneumococcal vaccine development (70). Similar assays have been employed in the assessment of functional immunity against *Campylobacter jejuni*, Group B *Streptococcus*, typhoidal and non-typhoidal *Salmonella* and *Neisseria gonorrhoeae* (71–75). While standardization of some of these pathogen-specific assays is pending, *de novo* design of *in vitro* correlates assessing functional activity of antibodies is not as straightforward for other pathogens, such as *Bordetella pertussis* (76). For facultative intracellular pathogens such as *Francisella tularensis* and MTB, for example, current correlate strategies in development are not suitable for assessments of mAb function as they involve peripheral blood lymphocytes only (77, 78). An added benefit of *in vitro* assessments of cloned mAb function, as a component of RV 2.0, is the needlessness of or significant reduction in usage of experimental animals. Efforts to develop and standardize *in vitro* correlates to assess mAb function are, hereby, merited.

Beyond bactericidal or opsonic functions, mAbs exhibit a variety of functions, including the modulation of cellular immune responses [extensively reviewed in Cooper (79), Amanna and Slika (80)], which require assessment. These functions also include toxin neutralization (useful in pertussis

and diphtheritic infections, for example) (81, 82) and increase in cellular cytotoxicity affecting intracellular pathogens such as *C. trachomatis* (83). Hence, non-bactericidal or non-opsonic mAbs, if exhibitive of these other functions, can still be utilized in other immunotherapeutic avenues.

CONCLUSION

With the increase in multidrug resistance among bacterial pathogens, the development of further effective preventive measures will be of significant benefit to public health. RV 2.0, a conceptually-advanced approach with the advantages of employing the natural host response (patient VH-VL combinations), relative speed, and reduction in animal use, has the potential to be a powerful tool in bacterial vaccine development. However, use of RV 2.0 is dependent on optimization of the technical aspects, and there are excellent prospects that this is achievable.

AUTHOR CONTRIBUTIONS

FB and PL conceptualization. FB, SS, and CG writing—original draft. FB and PL writing—reviewing and editing.

ACKNOWLEDGMENTS

We thank Dr. Mubarak Bidmos (Qatar University) and Dr. Victoria Wright (Imperial College London) for their critical but constructive review of the manuscript. PL and FB have received research grants from John and Michelle Bresnahan via MeningitisNow, and the Imperial College Confidence-in-Concept Award; and CG is funded by a BBSRC grant BB/R505742/1 from the National Productivity Investment Fund (NPIF), for work related to this manuscript.

REFERENCES

1. Calmette A. Preventive vaccination against tuberculosis with BCG. *Proc R Soc Med.* (1931) 24:85–94.
2. Salk JE, Lewis LJ, Bennett BL, Ward EN, Krech U, Youngner JS. Antigenic activity of poliomyelitis vaccines undergoing field test. *Am J Pub Health Nat Health* (1955) 45:151–62. doi: 10.2105/AJPH.45.2.151
3. Olin P, Rasmussen F, Gustafsson L, Hallander HO, Heijbel H. Randomised controlled trial of two-component, three-component, and five-component acellular pertussis vaccines compared with whole-cell pertussis vaccine. *Lancet* (1997) 350:1569–77. doi: 10.1016/S0140-6736(97)06508-2
4. Boslego J, Garcia J, Cruz C, Zollinger W, Brandt B, Ruiz S, et al. Efficacy, safety, and immunogenicity of a meningococcal group B (15: P1.3) outer membrane protein vaccine in Iquique, Chile. *Vaccine* (1995) 13:821–9. doi: 10.1016/0264-410X(94)00037-N
5. Galloway Y, Stehr-Green P, McNicholas A, O'Hallahan J. Use of an observational cohort study to estimate the effectiveness of the New Zealand group B meningococcal vaccine in children aged under 5 years. *Int J Epidemiol.* (2008) 38, 413–418. doi: 10.1093/ije/dyn228
6. Rosenqvist E, Høiby EA, Wedege E, Bryn K, Kolberg J, Klem A, et al. Human antibody responses to meningococcal outer membrane antigens after three doses of the Norwegian group B meningococcal vaccine. *Infect Immun.* (1995) 63:4642–52.
7. Robbins JB, Austrian R, Lee CJ, Rastogi SC, Schiffman G, Henrichsen J, et al. Considerations for formulating the second-generation pneumococcal capsular polysaccharide vaccine with emphasis on the cross-reactive types within groups. *J Infect Dis.* (1983) 148:1136–59. doi: 10.1093/infdis/148.6.1136
8. Schneerson R, Barrera O, Sutton A, Robbins JB. Preparation, characterization, and immunogenicity of *Haemophilus influenzae* type b polysaccharide-protein conjugates. *J Exp Med.* (1980) 152:361–76. doi: 10.1084/jem.152.2.361
9. Poolman JT, Peeters CC, Van Den Dobbelsteen GP. The history of pneumococcal conjugate vaccine development: dose selection. *Exp Rev Vaccines* (2013) 12:1379–94. doi: 10.1586/14760584.2013.852475
10. Anderson EL, Bowers T, Mink CM, Kennedy DJ, Belshe RB, Harakeh H, et al. Safety and immunogenicity of meningococcal A and C polysaccharide conjugate vaccine in adults. *Infect Immun.* (1994) 62:3391–5.
11. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* (1995) 269:496–512. doi: 10.1126/science.7542800
12. Pizza M, Scarlato V, Masignani V, Giuliani MM, Aricò B, Comanducci M, et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* (2000) 287:1816–20. doi: 10.1126/science.287.5459.1816
13. Mora M, Veggi D, Santini L, Pizza M, Rappuoli R. Reverse vaccinology. *Drug Discov. Today* (2003) 8:459–64. doi: 10.1016/S1359-6446(03)02689-8
14. Serruto D, Bottomley MJ, Ram S, Giuliani MM, Rappuoli R. The new multicomponent vaccine against meningococcal serogroup B, 4CMenB:

- immunological, functional and structural characterization of the antigens. *Vaccine* (2012) 30:B87–97. doi: 10.1016/j.vaccine.2012.01.033
15. Harris SL, Zhu D, Murphy E, McNeil LK, Wang X, Mayer LW, et al. Preclinical evidence for the potential of a bivalent fHBP vaccine to prevent *Neisseria meningitidis* serogroup C disease. *Hum Vaccin.* (2011) 7:68–74. doi: 10.4161/hv.7.0.14564
16. Mora M, Donati C, Medini D, Covacci A, Rappuoli R. Microbial genomes and vaccine design: refinements to the classical reverse vaccinology approach. *Curr Opin Microbiol.* (2006) 9:532–6. doi: 10.1016/j.mib.2006.07.003
17. Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, Scarselli M, et al. Identification of a universal Group B *Streptococcus* vaccine by multiple genome screen. *Science* (2005) 309:148–50. doi: 10.1126/science.1109869
18. Chiang MH, Sung WC, Lien SP, Chen YZ, Lo AFY, Huang JH, et al. Identification of novel vaccine candidates against *Acinetobacter baumannii* using reverse vaccinology. *Hum. Vaccin Immunother.* (2015) 11:1065–73. doi: 10.1080/21645515.2015.1010910
19. Meunier M, Guyard-Nicodème M, Hirschaud E, Parra A, Chemaly M, Dory D. Identification of novel vaccine candidates against *Campylobacter* through reverse vaccinology. *J Immunol Res.* (2016) 2016:5715790. doi: 10.1155/2016/5715790
20. Talukdar S, Zutshi S, Prashanth KS, Saikia KK, Kumar P. Identification of potential vaccine candidates against *Streptococcus pneumoniae* by reverse vaccinology approach. *Appl Biochem Biotechnol.* (2014) 172:3026–41. doi: 10.1007/s12010-014-0749-x
21. Hey A, Li MS, Hudson MJ, Langford PR, Kroll JS. Transcriptional profiling of *Neisseria meningitidis* interacting with human epithelial cells in a long-term *in vitro* colonization model. *Infect Immun.* (2013) 81:4149–59. doi: 10.1128/IAI.00397-13
22. Bittaye M, Cash P. *Streptococcus pneumoniae* proteomics: determinants of pathogenesis and vaccine development. *Exp Rev Proteomics* (2015) 12:607–21. doi: 10.1586/14789450.2015.1108844
23. Doro F, Liberatori S, Rodriguez-Ortega MJ, Rinaudo CD, Rosini R, Mora M, et al. Surrogate analysis as a fast track to vaccine discovery identification of a novel protective antigen for Group B *Streptococcus* hypervirulent strain COH1. *Mol Cell Proteomics* (2009) 8:1728–37. doi: 10.1074/mcp.M800486-MCP200
24. Burton DR. Antibodies, viruses and vaccines. *Nat Rev Immunol.* (2002) 2:706. doi: 10.1038/nri891
25. Rappuoli R, Bottomley MJ, D'Oro U, Finco O, De Gregorio E. Reverse vaccinology 2.0: human immunology instructs vaccine antigen design. *J Exp Med.* (2016) 213:469–81. doi: 10.1084/jem.20151960
26. Di Niro R, Mesin L, Zheng NY, Stamnaes J, Morrissey M, Lee JH, et al. High abundance of plasma cells secreting transglutaminase 2-specific IgA autoantibodies with limited somatic hypermutation in celiac disease intestinal lesions. *Nat Med.* (2012) 18:441–5. doi: 10.1038/nm.2656
27. Amara K, Steen J, Murray F, Morbach H, Fernandez-Rodriguez BM, Joshua V, et al. Monoclonal IgG antibodies generated from joint-derived B cells of RA patients have a strong bias toward citrullinated autoantigen recognition. *J Exp Med.* (2013) 210:445–55. doi: 10.1084/jem.20121486
28. Steinitz M, Klein G, Koskimies S, Makel O. EB virus-induced B lymphocyte cell lines producing specific antibody. *Nature* (1977) 269:420–22. doi: 10.1038/269420a0
29. Kozbor D, Roder JC, Chang TH, Steplewski Z, Koprowski H. Human anti-tetanus toxoid monoclonal antibody secreted by EBV-transformed human B cells fused with murine myeloma. *Hybridoma* (1982) 1:323–8. doi: 10.1089/hyb.1.1982.1.323
30. Winter G, Griffiths AD, Hawkins RE, Hoogenboom HR. Making antibodies by phage display technology. *Ann Rev Immunol.* (1994) 12:433–55. doi: 10.1146/annurev.iy.12.040194.002245
31. Cheung WC, Beausoleil SA, Zhang X, Sato S, Schieferl SM, Wieler JS, et al. A proteomics approach for the identification and cloning of monoclonal antibodies from serum. *Nat Biotechnol.* (2012) 30:447–52. doi: 10.1038/nbt.2167
32. Sato S, Beausoleil SA, Popova L, Beaudet JG, Ramenani RK, Zhang X, et al. Proteomics-directed cloning of circulating antiviral human monoclonal antibodies. *Nat Biotechnol.* (2012) 30:1039–43. doi: 10.1038/nbt.2406
33. DeKosky BJ, Ippolito GC, Deschner RP, Lavinder JJ, Wine Y, Rawlings BM, et al. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol.* (2013) 31:166–9. doi: 10.1038/nbt.2492
34. Smith K, Garman L, Wrammert J, Zheng NY, Capra JD, Ahmed R, et al. Rapid generation of fully human monoclonal antibodies specific to a vaccinating antigen. *Nat Protoc.* (2009) 4:372. doi: 10.1038/nprot.2009.3
35. Tiller T, Meffre E, Yurasov S, Tsuiji M, Nussenzweig MC, Wardemann H. Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *J Immunol Methods* (2008) 329:112–24. doi: 10.1016/j.jim.2007.09.017
36. Fink K. Origin and function of circulating plasmablasts during acute viral infections. *Front Immunol.* (2012) 3:78. doi: 10.3389/fimmu.2012.00078
37. Scheid JE, Mouquet H, Feldhahn N, Seaman MS, Velinzon K, Pietzsch J, et al. Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* (2009) 458:636–40. doi: 10.1038/nature07930
38. Muellenbeck MF, Ueberheide B, Amulic B, Epp A, Fenyo D, Busse CE, et al. Atypical and classical memory B cells produce *Plasmodium falciparum* neutralizing antibodies. *J Exp Med.* (2013) 210:389–99. doi: 10.1084/jem.20121970
39. Di Niro R, Mesin L, Raki M, Zheng NY, Lund-Johansen F, Lundin KE, et al. Rapid generation of rotavirus-specific human monoclonal antibodies from small-intestinal mucosa. *J Immunol.* (2010) 185:5377–83. doi: 10.4049/jimmunol.1001587
40. Pinder CL, Kratochvil S, Cizmeci D, Muir L, Guo Y, Shattock RJ, et al. Isolation and characterization of antigen-specific plasmablasts using a novel flow cytometry-based Ig capture assay. *J Immunol.* (2017) 199:4180–8. doi: 10.4049/jimmunol.1701253
41. Goodwin E, Gilman MS, Wrapp D, Chen M, Ngwuta JO, Moin SM, et al. Infants infected with respiratory syncytial virus generate potent neutralizing antibodies that lack somatic hypermutation. *Immunity* (2018) 48:339–49.e5. doi: 10.1016/j.immuni.2018.01.005
42. Lin Z, Chiang NY, Chai N, Seshasayee D, Lee WP, Balazs M, et al. *In vivo* antigen-driven plasmablast enrichment in combination with antigen-specific cell sorting to facilitate the isolation of rare monoclonal antibodies from human B cells. *Nat Protoc.* (2014) 9:1563–77. doi: 10.1038/nprot.2014.104
43. Ho IY, Bunker JJ, Erickson SA, Neu KE, Huang M, Cortese M, et al. Refined protocol for generating monoclonal antibodies from single human and murine B cells. *J Immunol Methods* (2016) 438:67–70. doi: 10.1016/j.jim.2016.09.001
44. Dodev TS, Karagiannis P, Gilbert AE, Josephs DH, Bowen H, James LK, et al. A tool kit for rapid cloning and expression of recombinant antibodies. *Sci Rep.* (2014) 4:5885. doi: 10.1038/srep05885
45. Blum LK, Adamska JZ, Martin DS, Rebman AW, Elliott SE, Cao RRL, et al. Robust B cell responses predict rapid resolution of Lyme disease. *Front Immunol.* (2018) 9:1634. doi: 10.3389/fimmu.2018.01634
46. Tan YC, Blum LK, Kongpachith S, Ju CH, Cai X, Lindstrom TM, et al. High-throughput sequencing of natively paired antibody chains provides evidence for original antigenic sin shaping the antibody response to influenza vaccination. *Clin Immunol.* (2014) 151:55–65. doi: 10.1016/j.clim.2013.12.008
47. Gupta P, Kamath AV, Park S, Chiu H, Lutman J, Maia M, et al. Preclinical pharmacokinetics of MHAA4549A, a human monoclonal antibody to influenza A virus, and the prediction of its efficacious clinical dose for the treatment of patients hospitalized with influenza A. *mAbs* 8:991–7. doi: 10.1080/19420862.2016.1167294
48. Mascola JR, D'Souza P, Gilbert P, Hahn BH, Haigwood NL, Morris L, et al. Recommendations for the design and use of standard virus panels to assess neutralizing antibody responses elicited by candidate Human Immunodeficiency Virus type 1 vaccines. *J Virol.* (2005) 79:10103–7. doi: 10.1128/JVI.79.16.10103-10107.2005
49. Borrow R, Aaberge IS, Santos GF, Eudey TL, Oster P, Glennie A, et al. Interlaboratory standardization of the measurement of serum bactericidal activity by using human complement against meningococcal serogroup B, strain 44/76-SL, before and after vaccination with the Norwegian MenBvac outer membrane vesicle vaccine. *Clin Diagn Lab Immunol.* (2005) 12:970–6. doi: 10.1128/CDLI.12.8.970-976.2005

50. Borrow R, Balmer P, Miller E. Meningococcal surrogates of protection — serum bactericidal antibody activity. *Vaccine* (2005) 23:2222–7. doi: 10.1016/j.vaccine.2005.01.051
51. Burton DR. What are the most powerful immunogen design vaccine strategies? Reverse vaccinology 2.0 shows great promise. *Cold Spring Harb Perspect Biol.* (2017) 9:a030262. doi: 10.1101/cshperspect.a030262
52. Kabanova A, Perez L, Lillier D, Marcandalli J, Agatic G, Becattini S, et al. Antibody-driven design of a human cytomegalovirus gHgLpUL128L subunit vaccine that selectively elicits potent neutralizing antibodies. *Proc Natl Acad Sci USA.* (2014) 111:17965–17970. doi: 10.1073/pnas.1415310111
53. Krarup A, Truan D, Furmanova-Hollenstein P, Bogaert L, Bouchier P, Bisschop IJ, et al. A highly stable prefusion RSV F vaccine derived from structural analysis of the fusion mechanism. *Nat Commun.* (2015) 6:8143. doi: 10.1038/ncomms9143
54. McLellan JS, Chen M, Leung S, Graepel KW, Du X, Yang Y, et al. Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* (2013) 340:1113–7. doi: 10.1126/science.1234914
55. Domachowske JB, Khan AA, Esser MT, Jensen K, Takas T, Villafana T, et al. Safety, tolerability, and pharmacokinetics of MEDI8897, an extended half-life single-dose respiratory syncytial virus prefusion F-targeting monoclonal antibody administered as a single dose to healthy preterm infants. *Pediatr Infect Dis J.* (2018) 37:886–92. doi: 10.1097/INF.00000000000001916
56. Nakamura G, Chai N, Park S, Chiang N, Lin Z, Chiu H, et al. An *in vivo* human-plasmablast enrichment technique allows rapid identification of therapeutic influenza A antibodies. *Cell Host Microb.* (2013) 14:93–103. doi: 10.1016/j.chom.2013.06.004
57. McBride JM, Lim JJ, Burgess T, Deng R, Derby MA, Maia M, et al. Phase 2 randomized trial of the safety and efficacy of MHAA4549A, a broadly neutralizing monoclonal antibody, in a human influenza A virus challenge model. *Antimicrob Agents Chemother.* (2017) 61:e01154–17. doi: 10.1128/AAC.01154-17
58. Lu DR, Tan YC, Kongpachith S, Cai X, Stein EA, Lindstrom TM, et al. Identifying functional anti-*Staphylococcus aureus* antibodies by sequencing antibody repertoires of patient plasmablasts. *Clin Immunol.* (2014) 152:77–89. doi: 10.1016/j.clim.2014.02.010
59. Zimmermann N, Thormann V, Hu B, Köhler AB, Imai-Matsushima A, Loch C, et al. Human isotype-dependent inhibitory antibody responses against *Mycobacterium tuberculosis*. *EMBO Mol Med.* (2016) 8:1325–39. doi: 10.15252/emmm.201606330
60. Olsen AW, Lorenzen EK, Rosenkrands I, Follmann F, Andersen P. Protective effect of vaccine promoted neutralizing antibodies against the intracellular pathogen *Chlamydia trachomatis*. *Front Immunol.* (2017) 8:1652. doi: 10.3389/fimmu.2017.01652
61. Bidmos FA, Nadel S, Srean GR, Kroll JS, Langford PR. Cross-reactive bactericidal anti-meningococcal antibodies can be isolated from convalescing invasive meningococcal disease patients using reverse vaccinology 2.0. *Front Immunol.* (2018) 9:1621. doi: 10.3389/fimmu.2018.01621
62. Dormitzer PR, Ulmer JB, Rappuoli R. Structure-based antigen design: a strategy for next generation vaccines. *Trends Biotechnol.* (2008) 26:659–67. doi: 10.1016/j.tibtech.2008.08.002
63. Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol.* (2014) 32:158–68. doi: 10.1038/nbt.2782
64. Hess JL, Blazer L, Romer T, Faber L, Buller RM, Boyle MD. Immunoproteomics. *J Chromatogr B* (2005) 815:65–75. doi: 10.1016/j.jchromb.2004.07.047
65. Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW, et al. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol.* (2012) 10:599–606. doi: 10.1038/nrmicro2850
66. Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnol J.* (2012) 7:620–34. doi: 10.1002/biot.201100155
67. Band VI, Ibegbu C, Kaur SP, Cagle SM, Tribble R, Jones CL, et al. Induction of human plasmablasts during infection with antibiotic-resistant nosocomial bacteria. *J Antimicrob Chemother.* (2014) 69:1830–3. doi: 10.1093/jac/dku047
68. Weitzner BD, Dunbrack RL Jr, Gray JJ. The origin of CDR H3 structural diversity. *Structure* (2015) 23:302–11. doi: 10.1016/j.str.2014.11.010
69. Mysara M, Leys N, Raes J, Monsieurs P. IPED: a highly efficient denoising tool for Illumina MiSeq Paired-end 16S rRNA gene amplicon sequencing data. *BMC Bioinformatics* (2016) 17:192. doi: 10.1186/s12859-016-1061-2
70. Song JY, Moseley MA, Burton RL, Nahm MH. Pneumococcal vaccine and opsonic pneumococcal antibody. *J Infect Chemother.* (2013) 19:412–25. doi: 10.1007/s10156-013-0601-1
71. Boyd MA, Tennant SM, Saag VA, Simon R, Muhsen K, Ramachandran G, et al. Serum bactericidal assays to evaluate typhoidal and nontyphoidal *Salmonella* vaccines. *Clin Vacc Immunol.* (2014) 21:712–21. doi: 10.1128/CI.00115-14
72. Le Doare K, Faal A, Jaiteh M, Sarfo F, Taylor S, Warburton F, et al. Association between functional antibody against Group B *Streptococcus* and maternal and infant colonization in a Gambian cohort. *Vaccine* (2017) 35:2970–8. doi: 10.1016/j.vaccine.2017.04.013
73. Necchi F, Saul A, Rondini S. Development of a high-throughput method to evaluate serum bactericidal activity using bacterial ATP measurement as survival readout. *PLoS ONE* (2017) 12:e0172163. doi: 10.1371/journal.pone.0172163
74. Piekarczyk A, Klyz A, Majchrzak M, Stein DC. Oral immunization of rabbits with *S. enterica* Typhimurium expressing *Neisseria gonorrhoeae* filamentous phage $\Phi 6$ induces bactericidal antibodies against *N. gonorrhoeae*. *Sci Rep.* (2016) 6:22549. doi: 10.1038/srep22549
75. Sahin O, Zhang Q, Meitzler JC, Harr BS, Morishita TY, Mohan R. Prevalence, antigenic specificity, and bactericidal activity of poultry anti-*Campylobacter* maternal antibodies. *Appl Environ Microbiol.* (2001) 67:3951–7. doi: 10.1128/AEM.67.9.3951-3957.2001
76. Brookes C, Freire-Martin I, Cavell B, Alexander F, Taylor S, Persaud R, et al. *Bordetella pertussis* isolates vary in their interactions with human complement components. *Emerg Microb Infect.* (2018) 7:81. doi: 10.1038/s41426-018-0084-3
77. De Pascalis R, Hahn A, Brook HM, Ryden P, Donart N, Mittereder L, et al. A panel of correlates predicts vaccine-induced protection of rats against respiratory challenge with virulent *Francisella tularensis*. *PLoS ONE* (2018) 13:e0198140. doi: 10.1371/journal.pone.0198140
78. Kurtz SL, Elkins KL. Correlates of vaccine-induced protection against *Mycobacterium tuberculosis* revealed in comparative analyses of lymphocyte populations. *Clin Vacc Immunol.* (2015) 22:1096–108. doi: 10.1128/CI.00301-15
79. Cooper AM. Cell-mediated immune responses in tuberculosis. *Ann Rev Immunol.* (2009) 27:393–422. doi: 10.1146/annurev.immunol.021908.132703
80. Amanna IJ, Slifka MK. Contributions of humoral and cellular immunity to vaccine-induced protection in humans. *Virology* (2011) 411:206–15. doi: 10.1016/j.virol.2010.12.016
81. Di Giovine P, Pinto A, Ölander RM, Sesardic D, Stickings P, Berbers G, et al. and Bujko, M. External quality assessment for the determination of diphtheria antitoxin in human serum. *Clin Vacc Immunol.* (2010) 17:1282–90. doi: 10.1128/CI.00096-10
82. Millen SH, Bernstein DI, Connelly B, Ward JI, Chang SJ, Weiss AA. Antibody-mediated neutralization of pertussis toxin-induced mitogenicity of human peripheral blood mononuclear cells. *Infect Immun.* (2004) 72:615–20. doi: 10.1128/IAI.72.1.615-620.2004
83. Moore T, Ekworomadu CO, Eko FO, MacMillan L, Ramey K, Ananaba GA, et al. Fc receptor-mediated antibody regulation of T cell immunity against intracellular pathogens. *J Infect Dis.* (2003) 188:617–24. doi: 10.1086/377134

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Bidmos, Siris, Gladstone and Langford. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Review on T Cell Epitopes Identified Using Prediction and Cell-Mediated Immune Models for *Mycobacterium tuberculosis* and *Bordetella pertussis*

Yuan Tian¹, Ricardo da Silva Antunes¹, John Sidney¹, Cecilia S. Lindestam Arlehamn¹, Alba Grifoni¹, Sandeep Kumar Dhanda¹, Sinu Paul¹, Bjoern Peters^{1,2}, Daniela Weiskopf¹ and Alessandro Sette^{1,2*}

¹ Division of Vaccine Discovery, La Jolla Institute for Immunology, La Jolla, CA, United States, ² Department of Medicine, University of California San Diego, La Jolla, CA, United States

OPEN ACCESS

Edited by:

Pedro A. Reche,
Complutense University of Madrid,
Spain

Reviewed by:

Pandjassame Kanguane,
Biomedical Informatics (P) Ltd., India
Etienne Caron,
Université de Montréal, Canada

*Correspondence:

Alessandro Sette
alex@lji.org

Specialty section:

This article was submitted to
Vaccines and Molecular Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 23 August 2018

Accepted: 12 November 2018

Published: 29 November 2018

Citation:

Tian Y, da Silva Antunes R, Sidney J, Lindestam Arlehamn CS, Grifoni A, Dhanda SK, Paul S, Peters B, Weiskopf D and Sette A (2018) A Review on T Cell Epitopes Identified Using Prediction and Cell-Mediated Immune Models for *Mycobacterium tuberculosis* and *Bordetella pertussis*. *Front. Immunol.* 9:2778. doi: 10.3389/fimmu.2018.02778

In the present review, we summarize work from our as well as other groups related to the characterization of bacterial T cell epitopes, with a specific focus on two important pathogens, namely, *Mycobacterium tuberculosis* (Mtb), the bacterium that causes tuberculosis (TB), and *Bordetella pertussis* (BP), the bacterium that causes whooping cough. Both bacteria and their associated diseases are of large societal significance. Although vaccines exist for both pathogens, their efficacy is incomplete. It is widely thought that defects and/or alteration in T cell compartments are associated with limited vaccine effectiveness. As discussed below, a full genome-wide map was performed in the case of Mtb. For BP, our focus has thus far been on the antigens contained in the acellular vaccine; a full genome-wide screen is in the planning stage. Nevertheless, the sum-total of the results in the two different bacterial systems allows us to exemplify approaches and techniques that we believe are generally applicable to the mapping and characterization of human immune responses to bacterial pathogens. Finally, we add, as a disclaimer, that this review by design is focused on the work produced by our laboratory as an illustration of approaches to the study of T cell responses to Mtb and BP, and is not meant to be comprehensive, nor to detract from the excellent work performed by many other groups.

Keywords: epitope, T cell, HLA, bacteria, *Mycobacterium tuberculosis*, *Bordetella pertussis*

EPITOPE IDENTIFICATION FOR MYCOBACTERIA

Our group has a long-standing interest in tuberculosis (TB). Early studies characterized a DR17-restricted epitope (1) and the effect of different T cell subsets on intracellular infection (2, 3). More recent studies characterized CD8 T cells (4). Subsequent studies described the first, to the best of our knowledge, true genome-wide screen for CD4 T cell epitopes (5). Subsequent studies report that CD4 T cells can recognize and target epitopes derived from mycobacterial ribosomal proteins and provide protective functions (6, 7).

The epitopes were further characterized by studies on the recognition of *Mycobacterium tuberculosis* (Mtb)-derived epitopes in a cohort with latent Mtb infection (8) and in diverse

populations from five continents (9) and intragenus conservation (10). We presented an analysis of the complexity of *Mtb*-specific epitopes in *Mtb* infected South Africans (11) and provided evidence that bi-allelic RORC mutations are detrimental to host immunity against *Mtb* (12). We further showed that transcriptomic analysis revealed novel immune signatures associated with TB (13–15) and the differentiation and function of T cells are influenced by the availability antigens (16).

In particular, previous studies (5) demonstrated the feasibility of utilizing genome-wide screen to identify human leukocyte antigen (HLA) class II epitopes derived from *Mtb*, based on combined bioinformatic predictions and high throughput *ex vivo* ELISPOT assays. Feasibility of the approach had previously been demonstrated for viral targets, however tackling a bacterial genome expressing over 4,000 open reading frames (ORFs) had not been attempted. Genome-wide screens have also been conducted to identify CD8 T cell *Mtb* epitopes (17–20). Notably, immunodominant CD8 T cell epitopes are enriched in cell wall and secreted proteins (18, 19). Future studies will utilize the same approach to focus on *Bordetella pertussis* (BP), which causes whooping cough.

EPITOPE IDENTIFICATION FOR OTHER BACTERIA ESPECIALLY BP

While our initial focus was mostly directed toward the study of *Mtb*, other microbes have also been studied. Maybeno et al. (21) described *Salmonella* epitopes, and Cannella et al. reported studies in *Brucella* (22). In the context of BP, we showed that initial whole-cell pertussis (wP) vaccination results in long-term Th1/Th17 polarization even with subsequent acellular boosters (23, 24).

It is hypothesized that the recent reemergence of BP infection is linked to the adoption of acellular pertussis (aP) vaccines based on specific BP antigens (FHA, Fim2/3, PRN, and PT). It is possible that the previous whole cell inactivated (wP) vaccine elicited a broader reactivity and targeted additional antigens, some of which might be of particular relevance and linked to superior vaccine performance. The extent and targets of T cell immunity in the context of natural infection and clinical disease are likewise not yet defined in a comprehensive fashion.

These considerations argue for performing broad epitope identification and characterization studies in BP as well. In the following sections we describe the techniques we have developed for the purpose of epitope identification and characterization, and then we describe specific applications to the TB and BP systems.

MEASURING HLA EPITOPE AFFINITY

Activation of alpha/beta classical T cells in general requires recognition of a specific peptide epitope, bound to specific major histocompatibility complex (MHC) molecules, a phenomenon classically named “HLA-restriction.” The methods used to establish restriction are described in a separate section below. Here we focus on the fact that, since HLA binding is a

prerequisite for a peptide being actually recognized as an epitope, measuring its HLA binding affinity is a powerful method to select epitope candidates. The relevant quantitative binding thresholds have been defined for both class I (25) and class II (26–28).

Our group has been a pioneer in the development of techniques to measure the binding of peptides to MHC molecules, termed HLA molecules in humans. Over the course of the last 30 years we have measured almost half a million MHC peptide binding constants for over 100,000 peptide/MHC combinations, and our group contributed a chapter describing our assay platform in detail to the laboratory compendium *Current Protocols in Immunology* (29).

The results obtained with this assay have been published in several 100 different peer reviewed journal articles. Our current assay panel allows measurements of binding to over 40 different HLA class I molecules and 35 HLA class II molecules. MHC binding is evaluated using a classical competition assay where peptides of interest competes with radiolabeled probe peptide for MHC binding (Figure 1). Plenty of supply of purified MHC molecules as well as labeled and unlabeled peptides is necessary for the establishment and usage of an MHC-peptide binding assay. Thus, our immunochemistry group has established an ongoing operation where cell lines expressing different alleles are expanded to allow for large-scale HLA purification by affinity chromatography.

Each cell line is rigorously characterized by HLA typing to ensure identity of the HLA allelic variant, and expression is monitored by flow cytometry. Standardized MHC purification protocols using affinity chromatography are in place. Furthermore, the purity and quantity of purified products are assessed by gel filtration and bicinchoninic acid assay (BCA) assays on each preparation. High affinity probes specific for each HLA have been identified for each allelic variant and used in a classic quantitative receptor ligand inhibition assay where IC₅₀ values are used to approximate true K_D values. Bound and unbound radiolabeled peptides are separated following incubation for 2 days, and their relative abundance is quantified. The specificity of each of the assays was rigorously determined by demonstrating high affinity binding of known independently defined eluted peptides or T cell epitopes restricted by the allelic variant in question.

HLA POLYMORPHISM AND BINDING PROMISCUITY

Thousands of different class I and class II HLA types exist in human populations. Most of the polymorphic residues line specific pockets in the HLA, which are involved in the peptide-HLA binding interaction. Accordingly, different HLA molecules have, in general, different binding specificities or “motifs” (30, 31). In fact, the capacity of general populations to bind a plethora of different sequences is the evolutionary force driving HLA polymorphism, to oppose the potential of pathogens to escape immune recognition by avoiding presentation of their peptides.

This extensive polymorphism poses a fundamental challenge for epitope identification and validation. A comprehensive

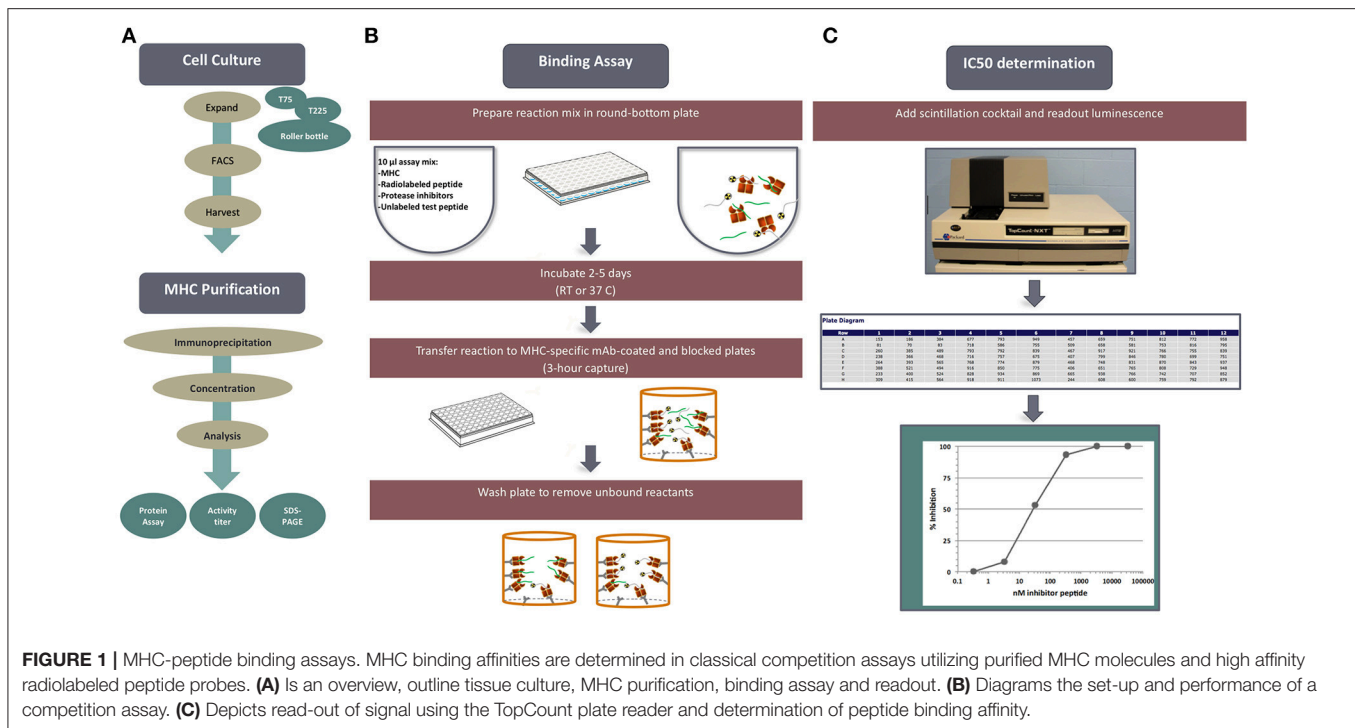


FIGURE 1 | MHC-peptide binding assays. MHC binding affinities are determined in classical competition assays utilizing purified MHC molecules and high affinity radiolabeled peptide probes. **(A)** Is an overview, outline tissue culture, MHC purification, binding assay and readout. **(B)** Diagrams the set-up and performance of a competition assay. **(C)** Depicts read-out of signal using the TopCount plate reader and determination of peptide binding affinity.

effort targeting thousands of variants would be impractical and unfeasible. A simple pragmatic solution would be to target HLA alleles that are present with the highest frequencies. However, this approach has to factor the reality that the frequencies of different HLA can vary dramatically across different ethnicities. Thus, comprehensive coverage of ethnically diverse populations requires careful analysis.

A phenomenon that counterbalances these difficulties and was originally recognized as a means to simplify the task is linked to HLA promiscuity. Indeed, it was already observed by the early 1990s that the same peptide could bind and be recognized by multiple HLA types (32–37), denoted as epitope “promiscuity.” Subsequent studies demonstrated that both classes I and II HLAs can be grouped into supertypes, defined on the basis of overlapping peptide binding repertoires (38–46). This is specifically applicable to HLA class II of the DR loci (33, 47, 48), and also DP (27, 49–51) and DQ (26, 51, 52).

A subsequent study used binding data of HLA DR, DQ, and DP to quantitatively assess how much promiscuity existed in HLA class II molecules. We found that these HLAs could be divided into seven major supertypes and, rather surprisingly, the repertoire overlap of class II supertypes was five to ten-fold higher than that of class I supertypes (51). These results indicated that if promiscuous binding would translate into promiscuous T cell recognition, then promiscuous epitopes might constitute a significant proportion of the total response.

PROMISCUOUS HLA CLASS II EPITOPES

In several independent studies and antigen systems we tested the hypotheses that promiscuous epitopes account for a significant

fraction of the specific immune response and that promiscuous epitopes can be identified by bioinformatic approaches. The notion that HLA class II promiscuous epitopes correspond to dominant epitopes accounting for a large fraction of the antigen specific response was initially evaluated with a set of overlapping 15-mer peptides spanning the Erythropoietin (EPO) protein (53). A large volume of subsequent data has demonstrated that these findings are generalizable to other systems, including proteins derived from infectious agents and allergens.

One series of experiments analyzed in detail the reactivity of allergic donors to the common allergen timothy grass (*Phleum pratense*), a mediator of hay fever (54). Over 40 different epitope regions were recognized, but upon closer inspection it was determined that only *nine* of them were required to cover 51% of the total response. These dominant regions were shown to correspond to promiscuously recognized epitopes, and shown to be predicted by bioinformatics algorithms targeting the most common DR, DP, and DQ variants.

This result was not limited to the timothy grass system. Indeed, similar results were obtained in several different allergen systems, including the *Blattella germanica* (Bla g) antigens associated with cockroach allergies (55). And, in a broader study, a panel of 133 allergens derived from 28 different sources, including fungi, trees, grasses, weeds, and indoor allergens, was surveyed utilizing predicted promiscuous HLA class II-binding peptides and ELISPOT assays with PBMC from allergic donors, resulting in the identification of 257 T cell epitopes (56).

In conclusion, a number of studies have shown that many peptides with highly promiscuous binding capacity are frequently recognized by immune individuals, and that promiscuous

recognition in the context of multiple HLA class II molecules may be a mechanism significantly contributing to epitope immunodominance (33, 53, 57–60). This might be related to the fact that promiscuous epitopes tend to bind HLA with high affinity, or simply that binding to multiple HLAs gives an epitope multiple contexts where it can be associated with immunogenicity. Several studies by our group have demonstrated that bioinformatic predictions directed toward selection of the most promiscuous binding peptides can identify a significant fraction of the pathogen or allergen specific response (54–56, 61).

DEVELOPMENT OF TOOLS TO PREDICT PROMISCUOUS EPITOPES

In the next series of investigations, we sought to derive and optimize a universal prediction schema based on data where sets of 15-mers overlapping by 10 amino acids representing the entire sequences of over 30 different allergens and bacterial proteins had been tested for T cell reactivity in human patients (62) (**Figure 2**). We specifically wanted to answer the question of how many predictions needed to be combined for maximal efficacy in real human patient populations, and furthermore, we wanted to determine which specific alleles should be included in such an optimal prediction tool. We defined optimal prediction parameters, and the resulting strategy was validated using a blind set of immunogenicity data that had not been utilized to derive the prediction scheme. We found that a 20th percentile IEDB consensus rank, combining predictions for a particular set of seven HLA class II, can predict about half of the total response. This approach can therefore be utilized as a universal prediction scheme, as it has been validated in a broad set of antigenic systems and in genetically diverse human patient populations.

In the study referenced above, the use of actual binding data instead of predictions did not improve the efficacy of the scheme, nor did performing allele specific predictions based on the particular HLA expressed in each individual patient. This indicated that, as expected the limited efficacy of the prediction was not due to the limitations of the algorithms, but rather that HLA binding predictions are associated with a rather high false positive rate, which fits well with the understanding that HLA binding is necessary but not sufficient for immunogenicity. Thus, other factors related to T cell repertoire and antigens processing also play a prominent role.

To address this issue, we used matched sets of dominant epitopes and negative peptides curated from the literature to train neural networks (63). The resulting “immunogenicity score” was further validated on 57 additional datasets (**Figure 2**). In all, data derived from more than 1,500 human donors and 2,000 peptides was considered in this training and validation effort. The results demonstrated that this agnostic “immunogenicity score” was effective in predicting dominant epitopes and human immunogenicity data. Surprisingly, the combination of immunogenicity score and HLA promiscuous predictions was associated with limited overall predictive improvement, suggesting, as previously noted, that antigen processing/T

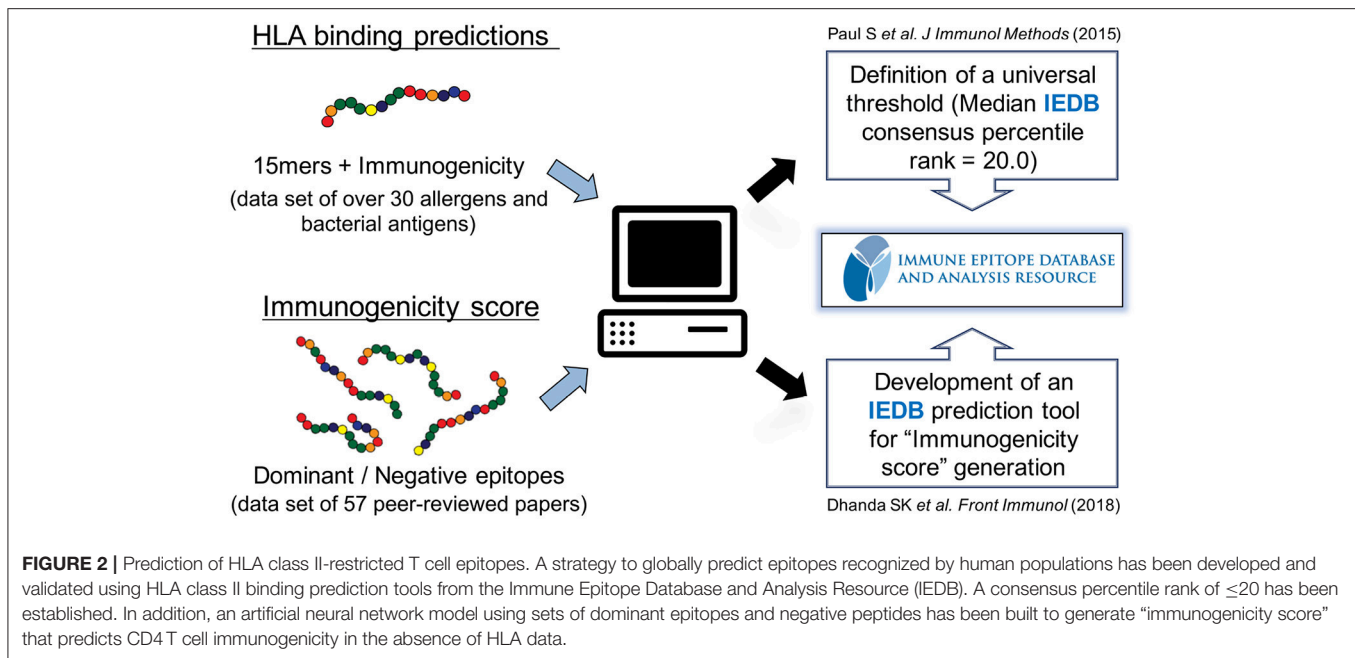
cell repertoire selection and HLA binding capacity might be influenced by coordinate evolution (64). Taken together, these results highlight that the bioinformatic tools necessary to identify promiscuous epitopes are available and have been validated in several independent studies, in different antigen system and different ethnicities.

DETERMINING HLA RESTRICTION OF T CELL RESPONSES

Determination of HLA restriction is a key element of epitope characterization, and precise knowledge of HLA restriction is also necessary to derive tetrameric staining reagents. HLA restriction was originally determined by the use of antibodies specific for different HLAs, in conjunction with antigen presentation assays; the essence of this strategy is to identify an antibody that blocks presentation of a given peptide to a given defined source of responding T cells (60). While straightforward in principle, this assay is often challenging, since antibodies with suitable specificity and selectivity are often not available, and T cells might promiscuously recognize the same peptide presented by multiple HLAs, yielding results difficult to interpret. Furthermore, the antibody and epitope concentration in the assay must be carefully controlled as excessive amounts of antibodies will inactivate the antigen-presenting cells (APCs) non-specifically, and high concentration of epitope will lead to self-presentation from the responding T cells.

An alternative is represented by determining whether the peptide can be presented by panels of partially HLA matched/mismatched cell lines/PBMCs. This is a powerful and simple approach, but can be limited by availability of suitable cell lines, and complicated again by promiscuous presentation, and the fact certain HLA combinations are in tight linkage disequilibrium. To overcome this limitation, we described an approach to define HLA class II restriction covering DP, DQ, and DR allelic variants that are most commonly represented in the general population (65). We specifically selected 46 DP, DQ, and DR HLAs which were projected to cover ~90% of these loci and constitute >66% of all the genes at each of these loci. Utilizing HLA data of actual populations from different geographical locations in the USA and Africa, we verified that these projections were accurate. A panel of single HLA transfected cell lines was developed and validated in a series of experiments, involving assessing HLA expression, identity, peptide binding, and epitope presentation (65).

The utility of this panel was further demonstrated by a quantitative study of HLA restriction and antigen-specific responses in a cohort of Mtb-immune individuals (11). Using APCs transfected with the panel of HLA class II molecules described above, HLA restrictions for nearly 300 different epitope/donor combinations were mapped. These results were the first large scale estimate of epitope complexity of CD4 T cell responses in a patient population and a microbial human pathogen, and indicated that the majority of epitopes were associated with promiscuous HLA restriction, further demonstrating the feasibility of the approach developed.



As an alternative complementary approach, we developed a method called Restrictor Analysis Tool for Epitopes (RATE) that can infer HLA restriction using CD4 T cell response data from HLA-typed individuals (66). The method, available online in the IEDB analysis resource, starts by inspecting, one epitope at a time, the HLA types present in individuals responding to each epitope. Then for each of these HLAs, calculates those enriched in frequency, comparing responders and non-responders to the specific epitope. The automated calculation yields a table of likely restrictions, Odds Ratios (ORs) and associated p -values. The method was validated by various experimental approaches, which derived strategies and thresholds for optimal performance (66, 67). The method is most effective for monogamous restrictions and by definition less able to detect promiscuous restrictions and HLA frequency variations due to genetic linkage.

ANALYSIS OF EPI TOPE CONSERVATION

Several lines of evidence indicate that sequence variability and conservation can have a dramatic effect on the shaping and effectiveness of immune responses in general and T cell responses in particular. This influence is dynamic and can have both positive and negative effects.

One broad series of effects relates to immunological pressure exerted by antimicrobial responses, of which probably amongst the most well noted cases are the widespread mutation of T cell epitopes observed in HIV and HCV (68, 69). It has to be underlined that pathogen escape by mutation is most effective for microbes with small genomes, or with responses of limited breadth, since simultaneous escape of a responses directed against large genomes and a large number of antigens/epitopes is by definition unlikely. It has indeed been proposed that the switch from wP to aP generated a response that is less diverse and

created an opportunity for BP to escape vaccine responses (70–72). It will be important to compare mutation rates of epitopes and non-epitopes, in wP, and aP antigens, to potentially either refute or support this hypothesis.

We have also noted that sequence variation or conservation can have a profound influence in shaping T cell responses by a different set of mechanisms. In general, we have noted that when individuals are exposed to different strains of the same species, or different species of phylogenetic related microbes, the immune response tends to focus on conserved epitopes. This is because repeated exposure of different but cross-reactive microbes ends up “teasing out” T cell recognizing conserved/homologous epitopes. Specifically, this has been observed in the case of dengue virus (DENV), where repeated exposure to different serotypes focuses the response on conserved epitopes (73, 74). The phenomenon is not limited to viruses, and is also observed in the case of grass pollens and ragweed pollen specific allergic responses (75, 76). In the case of bacterial genomes, we have shown that intragenus conservation within different mycobacteria species shapes T cell responses (10), and epitopes shared between mycobacteria tubercoid species and other non-pathogenic mycobacteria are preferentially recognized, indicating that differential reactivity may at least partially accounted for by environmental factors. It is currently unknown whether BP antigens and epitopes that share significant homology to other microbes encountered as a result of environmental exposure might be preferentially recognized.

In a separate study, we have recently shown that the sequence similarity between antigens and human microbiome can either dampen or increase T cell epitope immunogenicity (77). In this study, we systematically evaluated the homology of human microbiome sequences and sets of control peptides and T cell epitopes of various autoantigens, allergens, and infectious

pathogens. We expected that human adaptive immune system would be largely tolerant toward sequences identical or highly similar to those found in the human microbiome. We therefore predicted that these sequences would be more frequently found in the non-epitope category, as compared to the dominant epitope category. In many instances of epitope categories this was indeed the case, and reactivity was dampened (tolerogenic effect) suggesting that exposure to microbiome-derived sequence homologs might lead to T cell tolerization. However, in other cases, such as for example mycobacteria, and consistent with the studies mentioned above, the reactivity was increased (inflammatory effect) when the epitope sequence was conserved in the microbiome. It is currently unknown whether BP antigens and epitopes that share significant homology to other microbes contained in the human microbiome might be preferentially recognized or conversely tolerized.

VALIDATION AND CHARACTERIZATION OF T CELL EPITOPES

T cell epitopes can be characterized by various techniques such as mass spectrometry (MS), ELISPOT, intracellular cytokine staining (ICS), activation induced marker (AIM) assay, antigen-reactive T cell enrichment (ARTE) assay, tetramer staining, multidimensional fluorescence-based flow cytometry, and cytometry by time-of-flight (CyTOF), RNA-Sequencing (RNA-seq), and T cell receptor (TCR) sequencing (Figure 3). These techniques can also be combined. For example, performing TCR analysis of tetramer positive cells, or AIM/ARTE assays combined with ICS staining for particular cytokines.

In vitro vs. *ex vivo* Characterization

T cell responses can be characterized directly *ex vivo* or after *in vitro* re-stimulation in case that epitope-specific T cells are rare. Though *in vitro* re-stimulation allows for greater sensitivity, it may alter the phenotype of responding T cells; thus, the characterization of re-stimulated T cells require specific adjustments to the experimental strategy. Certain epitope characteristics are not altered by *in vitro* expansion, such as which particular TCR genes are expressed, HLA restriction, sequence conservation of the epitope recognized, or the pattern of cytokine polarization. On the other hand, memory and activation markers and other phenotypic markers usually detected by flow cytometry analysis are altered by the activation caused by cell culture. We have found that it is often possible to assess responses directly *ex vivo*, by using pools of different epitopes or peptides, so that the overall frequency of responding cells is enhanced. This approach is particularly effective when combined with the AIM assay described below, and particularly key to analyze samples with small volume. More specifically, our group has developed a megapool approach, which consists of large numbers of peptides (78). These “megapools” have been utilized in several systems such as allergies (79, 80), tuberculosis (11), tetanus and pertussis (24, 81), and DENV for both CD8 and CD4T cell epitopes (82–84).

Mass Spectrometry

MS-based approach has been utilized to identify and characterize T cell epitopes presented by MHC molecules since the 1990s (85, 86). Briefly, MHC molecules are purified from cell lysates and their associated peptides isolated and analyzed by MS. Although powerful and widely used, a full discussion of this approach is beyond the scope of the current review. Thus, we refer readers to (87) for more details on MS-based immunopeptidomics.

ELISPOT and ICS Assays

In our experience, ELISPOT assay is the most sensitive and high throughput-friendly method to measure T cell cytokine production. Our group has extensive experience using this method and we routinely utilize ELISPOT as a primary screen. In contrast, ICS assay is better at evaluating T cell phenotype and polyfunctionality. Both ELISPOT and ICS assays can characterize epitope pools even with small amounts of PBMCs. In our hands, we can characterize T cell responses with as little as 1 ml of peripheral blood.

AIM and ARTE Assays

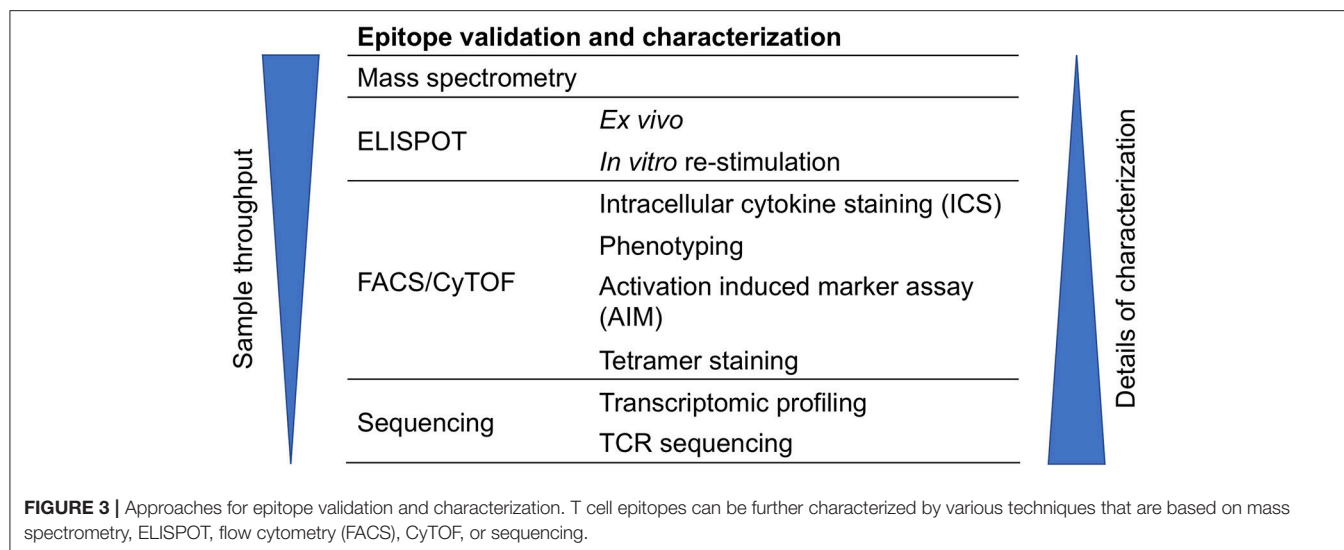
In addition to ICS, the selection of *ex vivo* activated antigen-specific CD4T cell populations can also be performed by measuring different activation molecules using the AIM assay [e.g., OX40 and CD25, which was co-developed by our group (81, 88)] or ARTE assay. The ARTE approach utilizes magnetic-enrichment of T cells that upregulate CD154 (CD40L) to assess human antigen-specific CD4T cells *ex vivo* (89). ARTE has been applied to identify antigen-specific T cells for several infections and could select rare antigen-specific T cells after short stimulation period without the need for ICS (90).

Tetramer Staining

This approach identifies antigen-specific T cells using tetramer staining reagents (91, 92). Furthermore, tetramer enrichment technique can be utilized if the frequency of antigen-specific T cells is low (93). However, specific reagents for each unique HLA:epitope combination of interest must be produced in order to use this approach. Thus, it is usually used for in-depth characterization of T cells of selected epitope-specificities and HLA restrictions.

Multidimensional Flow Cytometry and CyTOF

These are powerful techniques to characterize cell samples by evaluating the expression of many different markers associated with cell lineages (94), activation and functional activities (95), memory cell subtypes and chemokine receptor expression (96). Multicolor fluorescence-based flow cytometry is in general more user and equipment friendly, as antibodies are more generally available. In addition, this technique allows recovery of the cells by cell sorting, and thus is readily coupled with transcriptomic analysis. In contrast to flow cytometry, CyTOF can detect, discriminate, and quantify antibodies that are conjugated to various heavy-metal isotopes with high accuracy (97). This avoids spectral overlap between fluorophores and



allows measuring more cellular parameters simultaneously. High-dimensional phenotypic data can be visualized using algorithms such as visualization of stochastic neighbor embed (viSNE) and spanning-tree progression analysis of density-normalized events (SPADE) (98). We have utilized CyTOF and viSNE to visualize and characterize the heterogeneity of human CD4 effector memory T re-expressing CD45RA (Temra) cells (95).

Transcriptomic Profiling

Epitope-specific T cells can be further characterized in-depth by transcriptomic profiling that uses deep-sequencing technologies, including bulk RNA-seq or single-cell RNA-seq (scRNA-seq). By comparison with bulk RNA-seq, scRNA-seq is a more powerful tool to address cellular heterogeneity and to identify novel subpopulations in a “hypothesis free” manner, since individual cells within the “same” population may differ dramatically (99–101). Gene expression profiling using these methodologies is routinely utilized in our laboratory. Examples include the definition of signatures predictive of latent tuberculosis infection (13), the characterization of CD4 cytotoxic memory T cells (95, 102, 103) or CD4 differential responses to BP primary vaccination after aP boost vaccination (23).

TCR Sequencing

In addition to functional and phenotypic characterization of epitope-specific T cell responses, one can further define their TCR repertoires by TCR sequencing (104). TCRs dictate the antigen specificity of T cells through the interactions with peptide and major histocompatibility complexes. By analyzing epitope-associated TCR repertoires, it is possible to investigate common features of TCRs that are specific for a particular epitope and identify determinants that may predict specificity (105, 106). Thus, this strategy will enable researchers to systematically integrate epitopes with their specific TCR sequences as well as their associated T cell responses.

GENOME-WIDE SCREEN OF *Mtb* HLA CLASS II EPITOPES

As a way to illustrate how the various techniques can be utilized to tackle even large complex microbial genomes, we briefly summarize the results of an *Mtb* genome-wide screen (5). Our general strategy has been to first study in detail a limited number of well characterized dominant antigens, to investigate the mechanisms associated with immunodominance, and provide a point of reference for the genome-wide screen (60). While several dominant antigens were known and well described, a truly systematic and unbiased screen had not been attempted before, due to the complexity of the genome and the large number of ORFs. Next, as summarized below, we performed an unbiased genome-wide screen, and based on the results we selected the dominant epitopes and antigens (9). These were then utilized to characterize the epitope and the phenotype of the associated T cells (8, 10, 11, 11, 14), and also to develop a *Mtb* epitope megapool that was utilized in numerous studies and has proven a valuable tool to analyzed responses in a number of different settings (8, 11, 13, 105).

To perform a genome-wide screen of *Mtb*, we selected all full genome sequences available at that point, and utilized the approaches described above, to define a library of about 20,000 predicted promiscuous binders (5). These were synthesized, and screened first as pools and then in deconvolution experiments to identify the actual epitopes responsible for T cell activation. The library also included over 1,500 different variants not totally conserved amongst the genomes analyzed. Here it could be noted that the capacity to readily test for sequence variants is an advantage of our approach.

We have identified hundreds of different epitopes; the response was thus remarkably broad, and each individual recognized tens of different epitopes, the dominant epitopes, and antigens varied appreciably from one individual to the next. The overwhelming majority of the response was CD4 restricted, which was not unexpected since the epitopes were

identified based on their predicted ability to bind to HLA class II molecules. When the epitopes were mapped back to their antigen of origin using the H37Rv reference genome, a set of 82 antigens were identified as dominant, in that they accounted for about 80% of the total response. The majority of these antigens were not previously identified as T cell antigens.

Further analysis revealed that the vast majority of the response mapped to very discrete regions of the *Mtb* genome, and specifically to three clusters of reactivity within the genome, which encoded close to half of the total reactivity. One of the islands contained the well-characterized antigens early secretory antigenic target 6 kDa (ESAT-6) and culture filtrate protein 10 kDa (CFP10), secreted by Type VII secretion systems (T7SS or Esx system). The other two islands also contained Type VII secretion protein pairs. To further highlight the novelty of these observation, we discovered that the antigens that were recognized as dominant were not limited to the secreted proteins, but also included proteins from the actual secretion apparatus. Thus, the results obtained illustrated the feasibility of the approach, while at the same time identifying a number of novel epitopes and antigens, and providing new insights into the mechanisms of immunodominance.

THE RESURGENCE OF BP AS A PUBLIC HEALTH CONCERN

BP has been a health concern since the Middle Ages (107), and whooping cough was prevalent and associated with high morbidity and mortality until the widespread vaccination (108). Vaccination with wP vaccine in general population has greatly reduced whooping cough since the 1950s. Nevertheless, the wP vaccine was associated with of minor adverse reactions and very rare serious side-effects, which resulted in its replacement by the aP vaccine in the United States (109, 110). In spite of widespread vaccination, the cases of whooping cough have recently been steadily increasing in the United States (www.cdc.gov). Epidemiological evidence indicates that the increased prevalence may be associated with the switch from wP to aP vaccine in the mid-1990s, further implicating a potential role for waning immunity (www.cdc.gov).

Although the phenomenon of “waning BP immunity” is a serious issue (111), it is not straightforward to address as its manifestation appears more than 15 years following the initial vaccination. Therefore, it is crucial to understand the underlying mechanisms of waning immunity in order to guide the design of effective vaccines. In addition to qualitative differences in the response, several other mechanisms may exist. Two main additional hypotheses have been put forth (Figure 4). First, as the wP vaccine contains >3,400 ORFs, whereas the aP vaccine includes only a few BP proteins, it is likely that a differential breadth of response is induced by the wP and aP vaccines (112). Furthermore, the chemically detoxified pertussis toxin (PT) contained in the aP vaccine may have altered antigenicity and could potentially influence vaccine efficacy (113, 114). Second, it has been proposed that

decreased vaccine efficacy might be due to antigenic drift (108, 115–119).

Both antibody and T cell responses are thought to be associated with the effectiveness of pertussis vaccination. Notably, protective immunity against BP persists even after antibody levels have reduced (120–122), suggesting that T cells play a role in long-term protection against BP. Animal studies suggest that memory CD4 T cells of Th1 and Th17 phenotype mediate for long-term protection, which are induced by infection as well as wP vaccination (123–125). In contrast, aP vaccination is associated with a predominant Th2 response in humans (126–129). Furthermore, a few studies have reported that aP vaccination induces qualitative changes in T cell responses, resulting suboptimal efficacy (130–133) (Figure 4).

GENOME-WIDE SCREEN OF T CELL RESPONSE TO BP

To date, the question of whether the wP vaccine elicits strong T cell responses to additional and different set of antigens to those elicited by aP vaccination, and if so which antigens, has not been addressed. Given the fact that our genome-wide screen of *Mtb* (5) revealed novel dominant antigens that had escaped detection, despite decades of investigation of *Mtb*-specific T cell responses, we consider this possibility likely. By the same token, the breadth of responses induced by natural infection and clinical disease are not known. Here as well, it is likely that additional antigens beyond those included in the current vaccine are of importance; for example, the ACT toxin has been shown to be targeted by BP infected individuals, and the combination of PT and ACT results in superior protection from disease in animal models of BP infection and disease (134–136). These considerations underscore further investigation of BP antigens and T cell epitopes, as well as correlates of protection.

DEFINITION OF aP EPITOPES FOLLOWING aP vs. wP VACCINATION

We previously completed a series of studies aimed at the definition of CD4 T cell epitopes derived from the antigens contained in the aP vaccine (24). These illustrate the general feasibility of the study of epitopes and T cell reactivity in BP, and also provide a point of reference to interpret the results obtained in a potential genomic screen of BP T cell reactivity.

In those studies (24), PBMCs from either aP- or wP-primed healthy adults with recent aP booster were used to screen overlapping peptides derived from the protein components that are the foundation of the acellular vaccine: PT, Pertactin, Filamentous hemagglutinin, and Fimbrae 2 & 3). We utilized high-throughput *ex vivo* ELISPOT assays to measure T cell cytokine production of interferon- γ (IFN- γ) and IL-5, and deconvolution of positive peptide pools identified individual T cell epitopes. Epitope mapping revealed the same epitopes were recognized by both aP- and wP-primed individuals (24). However, the ratios of IFN- γ and IL-5 revealed a Th1 bias in originally wP-primed donors, and

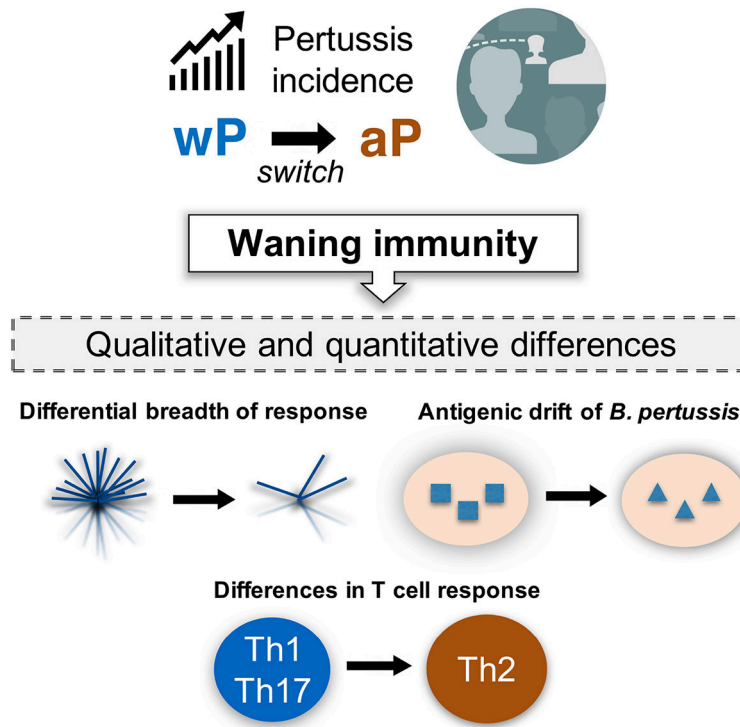


FIGURE 4 | Incidence of pertussis and proposed models of waning immunity. The phenomenon of resurgence of pertussis is gradually increasing as a public health concern, even in countries with high vaccination coverage. It would be important to define the mechanisms associated with waning immunity based on our current knowledge of the qualitative and quantitative changes in both T cell response and BP genetic evolution under vaccine pressure.

dominance of IL-5 in individuals primed with aP (24). This differential polarization persists following booster, even decades after original priming (24).

CHARACTERIZATION AND VALIDATION OF EPITOPES DERIVED FROM aP-ANTIGENS

As a result of the studies described in the previous section, we defined a “megapool” encompassing the 132 most dominant epitopes recognized, which allowed to assess BP responses directly *ex vivo* using the AIM assay combined with ICS assays, without need for *in vitro* re-stimulation, and thus allow direct phenotyping avoid the alterations induced by the *in vitro* re-stimulation step. This strategy was utilized to evaluate the phenotype and function of T cells in the PBMCs from either wP- or aP-primed donors, following an aP booster 1-3 months post-vaccination, to allow for memory T cells return to steady state conditions (23).

Using the *ex vivo* readouts we still detected the persistent differential polarization previously detected after *in vitro* re-stimulation. Moreover, we detected differential polarization toward IL-4 and IL-9 in aP-primed donors and IFN- γ and IL-17 in wP-primed donors (23). This effect was specific for the vaccine antigens, since no difference was noted for other epitopes such as megapools from the ubiquitous antigens CMV and EBV. The IL-17 polarization of wP vaccination had been previously in baboon models, but not for humans (124, 125, 137, 138).

The observation of IL-9 differential polarization is a novel aspect of our study. In-depth phenotypic analysis using combined ICS and transcriptomic analysis of BP-specific memory T cells from aP vs. wP donors revealed clear differences, especially at the level of effector memory T (Tem) cell. 13 differentially expressed genes were identified by comparing aP-Tem and wP-Tem cells, including *IL9* and *TGIF2*, which is related to regulation of TGF- β -responsive genes (139, 140). *IL5*, *IL13*, and *TGFB1* were also up-regulated in samples from aP donors (23).

In contrast to aP prime, wP prime is associated with substantially higher magnitude of CD4 T cell responses following aP booster, when *ex vivo* responses were assayed in a time window ranging from a few days to several months. Consistent with these findings, by the use of *in vitro* proliferation assays we could show that the aP originally primed donors were associated with lower proliferative capacity (23). In conclusion, these results demonstrate that the various techniques described above can be used to dissect and define the phenotype associated with BP specific T cell responses, and reveal important biological differences.

CONSERVATION OF BP EPITOPES ACROSS BP VARIANTS

As mentioned before, previous studies (108, 115–119) indicated that mutation might be accumulating in the acellular vaccine antigens, and that this phenomenon might be related to

the apparent waning of BP immunity. Conversely, we also have recently shown that the human microbiome composition modulates T cell responses via molecular mimicry (141).

A first line of preliminary analysis, considers the possibility that new BP strains that carry mutations at key epitopes (pathogen escape), have evolved. Several studies have identified mutations in circulating BP strains that could be the result of pathogen adaptation to immune pressure. For example, Bart et al. (115) identified a total of 471 coding SNPs (genetic variations that result in amino acid changes in the encoded proteins) from pre-vaccination strains. Precise mapping of the T cell epitopes that are prevalently recognized in the human population, including for antigens that are not also targets of immune responses, will further elucidate if the observed genetic variability of circulating BP strains is indeed a result of T cell immune pressure.

CONCLUSION AND PERSPECTIVE

There is strong evidence suggesting that T cells have important functions in BP immunity and vaccine efficacy. However, T cell epitopes elicited by either natural infection or whole cell (wP) vaccination have not been comprehensively defined, and the corresponding T cell phenotypes have not been characterized. Based upon the success of genome-wide screen of *Mtb*, we make an argument to support performing a genome-wide screen of T cell responses in individuals vaccinated with wP vaccines, and individuals previously diagnosed with whooping cough disease, to understand the targets of cellular immunity in those conditions. Such an investigation could utilize techniques developed and validated over the years, which include both direct *ex vivo* assays such as the AIM assay and *in vitro* expansion of memory T cells utilizing BP lysates. Although powerful, the full genome-wide screen approach also has its limitations. For instance, this approach will not select non-canonical peptides presented by HLA molecules such as peptides originating from non-coding regions and spliced peptides. In fact, recent studies (one of which we coauthored) indicate that a substantial fraction of the HLA peptidome (class I but probably

class II as well) is composed of hybrid peptides that originate from two different peptide fragments (so-called cis-spliced or trans-spliced peptides) (142, 143). If general rules that predict the splicing mechanisms can be defined, these sliced peptides could be predicted and thus incorporated in the analysis.

T cell responses against the various epitopes and associated antigens can be characterized and validated using several different complementary approaches. These include determining HLA restriction, and measuring HLA binding affinity, characterizing memory phenotypes, functionality and helper T cell subsets, and patterns of epitope sequence variation. Additionally, it would be of considerable interest to characterize transcriptomic profiles associated with recognition of the new epitopes identified, as compared to the ones currently included in the aP vaccine. These studies could potentially address several hypotheses proposed to explain the decreased efficacy of aP vaccines, namely differences in antigen specificity, differences in functionality, and/or mutations associated with the antigen/epitope associated with vaccine responses. Furthermore, we anticipate that these studies will be broadly applicable to other intracellular bacterial pathogens such as *Salmonella* and *Brucella*.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by National Institutes of Health contracts and grants HHSN272200900044C, HHSN272201200010C, HHSN272201400045C, U19 AI118626, and U01 AI141995.

ACKNOWLEDGMENTS

We thank Eugene L. Moore for designing **Figure 1**.

REFERENCES

- Geluk A, van Meijgaarden KE, de Vries RR, Sette A, Ottenhoff TH. A DR17-restricted T cell epitope from a secreted *Mycobacterium tuberculosis* antigen only binds to DR17 molecules at neutral pH. *Eur J Immunol.* (1997) 27:842–7. doi: 10.1002/eji.1830270406
- Stenger S, Mazzaccaro RJ, Uyemura K, Cho S, Barnes PF, Rosat JP, et al. Differential effects of cytolytic T cell subsets on intracellular infection. *Science* (1997) 276:1684–7. doi: 10.1126/science.276.5319.1684
- Cho S, Mehra V, Thoma-Uszynski S, Stenger S, Serbina N, Mazzaccaro RJ, et al. Antimicrobial activity of MHC class I-restricted CD8+ T cells in human tuberculosis. *Proc Natl Acad Sci USA* (2000) 97:12210–5. doi: 10.1073/pnas.210391497
- Lewinsohn DA, Winata E, Swarbrick GM, Tanner KE, Cook MS, Null MD, et al. Immunodominant tuberculosis CD8 antigens preferentially restricted by HLA-B. *PLoS Pathog.* (2007) 3:1240–9. doi: 10.1371/journal.ppat.0030127
- Lindestam Arlehamn CS, Gerasimova A, Mele F, Henderson R, Swann J, Greenbaum JA, et al. Memory T cells in latent *Mycobacterium tuberculosis* infection are directed against three antigenic islands and largely contained in a CXCR3+CCR6+ Th1 subset. *PLoS Pathog.* (2013) 9:e1003130. doi: 10.1371/journal.ppat.1003130
- Kennedy SC, Johnson AJ, Bharrhan S, Lindestam Arlehamn CS, Xu J, Garforth SJ, et al. Identification of mycobacterial ribosomal proteins as targets for CD4(+) T cells that enhance protective immunity in tuberculosis. *Infect Immun.* (2018). 86:e00009–18. doi: 10.1128/IAI.00009-18
- Johnson AJ, Kennedy SC, Lindestam Arlehamn CS, Goldberg MF, Saini NK, Xu J, et al. Identification of mycobacterial RplJ/L10 and RpsA/S1 proteins as novel targets for CD4(+) T Cells. *Infect Immun.* (2017) 85:e01023–16. doi: 10.1128/IAI.01023-16
- Scriba TJ, Carpenter C, Pro SC, Sidney J, Musvosvi M, Rozot V, et al. Differential recognition of *Mycobacterium tuberculosis*-specific epitopes as a function of tuberculosis disease history. *Am J Respir Crit Care Med.* (2017) 196:772–81. doi: 10.1164/rccm.201706-1208OC
- Carpenter C, Sidney J, Kolla R, Nayak K, Tomiyama H, Tomiyama C, et al. A side-by-side comparison of T cell reactivity to fifty-nine *Mycobacterium tuberculosis* antigens in diverse populations from five continents. *Tuberculosis* (2015) 95:713–21. doi: 10.1016/j.tube.2015.07.001

10. Lindestam Arlehamn CS, Paul S, Mele F, Huang C, Greenbaum JA, Vita R, et al. Immunological consequences of intragenus conservation of *Mycobacterium tuberculosis* T-cell epitopes. *Proc Natl Acad Sci USA*. (2015) 112:E147–55. doi: 10.1073/pnas.1416537112
11. Lindestam Arlehamn CS, McKinney DM, Carpenter C, Paul S, Rozot V, Makgotlho E, et al. A quantitative analysis of complexity of human pathogen-specific CD4 T cell responses in healthy *M. tuberculosis* Infected South Africans. *PLoS Pathog*. (2016) 12:e1005760. doi: 10.1371/journal.ppat.1005760
12. Okada S, Markle JG, Deenick EK, Mele F, Averbuch D, Lagos M, et al. IMMUNODEFICIENCIES. Impairment of immunity to *Candida* and *Mycobacterium* in humans with bi-allelic RORC mutations. *Science* (2015) 349:606–13. doi: 10.1126/science.aaa4282
13. Burel JG, Lindestam Arlehamn CS, Khan N, Seumois G, Greenbaum JA, Taplitz R, et al. Transcriptomic analysis of CD4(+) T cells reveals novel immune signatures of latent tuberculosis. *J Immunol*. (2018) 200:3283–90. doi: 10.4049/jimmunol.1800118
14. Arlehamn CL, Seumois G, Gerasimova A, Huang C, Fu Z, Yue X, et al. Transcriptional profile of tuberculosis antigen-specific T cells reveals novel multifunctional features. *J Immunol*. (2014) 193:2931–40. doi: 10.4049/jimmunol.1401151
15. Lindestam Arlehamn CS, Sette A. Definition of CD4 immunosignatures associated with MTB. *Front Immunol*. (2014) 5:124. doi: 10.3389/fimmu.2014.00124
16. Moguche AO, Musvosvi M, Penn-Nicholson A, Plumlee CR, Mearns H, Geldenhuys H, et al. Antigen availability shapes T cell differentiation and function during tuberculosis. *Cell Host Microbe* (2017) 21:695–706 e5. doi: 10.1016/j.chom.2017.05.012
17. Caccamo N, Guggino G, Meraviglia S, Gelsomino G, Di Carlo P, Titone L, et al. Analysis of *Mycobacterium tuberculosis*-specific CD8 T-cells in patients with active tuberculosis and in individuals with latent infection. *PLoS ONE* (2009) 4:e5528. doi: 10.1371/journal.pone.0005528
18. Tang ST, van Meijgaarden KE, Caccamo N, Guggino G, Klein MR, van Weeren P, et al. Genome-based *in silico* identification of new *Mycobacterium tuberculosis* antigens activating polyfunctional CD8+ T cells in human tuberculosis. *J Immunol*. (2011) 186:1068–80. doi: 10.4049/jimmunol.1002212
19. Lewinsohn DA, Swarbrick GM, Park B, Cansler ME, Null MD, Toren KG, et al. Comprehensive definition of human immunodominant CD8 antigens in tuberculosis. *NPJ Vaccines* (2017) 2:11. doi: 10.1038/s41541-017-0008-6
20. Lindestam Arlehamn CS, Lewinsohn D, Sette A, Lewinsohn D. Antigens for CD4 and CD8 T cells in tuberculosis. *Cold Spring Harb Perspect Med*. (2014) 4:a018465. doi: 10.1101/cshperspect.a018465
21. Maybeno M, Redeker A, Welten SP, Peters B, Loughhead SM, Schoenberger SP, et al. Polyfunctional CD4+ T cell responses to immunodominant epitopes correlate with disease activity of virulent *Salmonella*. *PLoS ONE* (2012) 7:e43481. doi: 10.1371/journal.pone.0043481
22. Cannella AP, Tsolis RM, Liang L, Felgner PL, Saito M, Sette A, et al. Antigen-specific acquired immunity in human brucellosis: implications for diagnosis, prognosis, and vaccine development. *Front Cell Infect Microbiol*. (2012) 2:1. doi: 10.3389/fcimb.2012.00001
23. da Silva Antunes R, Babor M, Carpenter C, Khalil N, Cortese M, Mentzer AJ, et al. Th1/Th17 polarization persists following whole-cell pertussis vaccination despite repeated acellular boosters. *J Clin Invest*. (2018) 128:3853–65. doi: 10.1172/JCI121309
24. Bancroft T, Dillon MB, da Silva Antunes R, Paul S, Peters B, Crotty S, et al. Th1 versus Th2 T cell polarization by whole-cell and acellular childhood pertussis vaccines persists upon re-immunization in adolescence and adulthood. *Cell Immunol*. (2016) 304–305:35–43. doi: 10.1016/j.cellimm.2016.05.002
25. Sette A, Vitiello A, Rehman B, Fowler P, Nayarsina R, Kast WM, et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J Immunol*. (1994) 153:5586–92.
26. Sidney J, Steen A, Moore C, Ngo S, Chung J, Peters B, et al. Divergent motifs but overlapping binding repertoires of six HLA-DQ molecules frequently expressed in the worldwide human population. *J Immunol*. (2010) 185:4189–98. doi: 10.4049/jimmunol.1001006
27. Sidney J, Steen A, Moore C, Ngo S, Chung J, Peters B, et al. Five HLA-DP molecules frequently expressed in the worldwide human population share a common HLA supertypic binding specificity. *J Immunol*. (2010) 184:2492–503. doi: 10.4049/jimmunol.0903655
28. Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, et al. Several common HLA-DR types share largely overlapping peptide binding repertoires. *J Immunol*. (1998) 160:3363–73.
29. Sidney J, Southwood S, Moore C, Oseroff C, Pinilla C, Grey HM, et al. Measurement of MHC/peptide interactions by gel filtration or monoclonal antibody capture. *Curr Protoc Immunol*. (2013) Chapter 18:Unit 18 3. doi: 10.1002/0471142735.im1803s100
30. Madden DR. The three-dimensional structure of peptide-MHC complexes. *Annu Rev Immunol*. (1995) 13:587–622. doi: 10.1146/annurev.iy.13.040195.003103
31. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* (1999) 50:213–9. doi: 10.1007/s002510050595
32. Sinigaglia F, Guttinger M, Kilgus J, Doran DM, Matile H, Etlinger H, et al. A malaria T-cell epitope recognized in association with most mouse and human MHC class II molecules. *Nature* (1988) 336:778–80. doi: 10.1038/336778a0
33. Alexander J, Sidney J, Southwood S, Ruppert J, Oseroff C, Maewal A, et al. Development of high potency universal DR-restricted helper epitopes by modification of high affinity DR-blocking peptides. *Immunity* (1994) 1:751–61. doi: 10.1016/S1074-7613(94)80017-0
34. O'Sullivan D, Arrhenius T, Sidney J, Del Guercio MF, Albertson M, Wall M, et al. On the interaction of promiscuous antigenic peptides with different DR alleles. Identification of common structural motifs. *J Immunol*. (1991) 147:2663–9.
35. Rothbard JB, Lechler RI, Howland K, Bal V, Eckels DD, Sekaly R, et al. Structural model of HLA-DR1 restricted T cell antigen recognition. *Cell* (1988) 52:515–23. doi: 10.1016/0092-8674(88)90464-3
36. Krieger JI, Karr RW, Grey HM, Yu WY, O'Sullivan D, Batovsky L, et al. Single amino acid changes in DR and antigen define residues critical for peptide-MHC binding and T cell recognition. *J Immunol*. (1991) 146:2331–40.
37. Panina-Bordignon P, Tan A, Termijtellen A, Demotz S, Corradin G, Lanzavecchia A. Universally immunogenic T cell epitopes: promiscuous binding to human MHC class II and promiscuous recognition by T cells. *Eur J Immunol*. (1989) 19:2237–42. doi: 10.1002/eji.1830191209
38. Sette A, Sidney J. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* (1999) 50:201–12. doi: 10.1007/s002510050594
39. Sette A, Sidney J. HLA supertypes and supermotifs: a functional perspective on HLA polymorphism. *Curr Opin Immunol*. (1998) 10:478–82. doi: 10.1016/S0952-7915(98)80124-6
40. Sidney J, del Guercio MF, Southwood S, Engelhard VH, Appella E, Rammensee HG, et al. Several HLA alleles share overlapping peptide specificities. *J Immunol*. (1995) 154:247–59.
41. Sidney J, Grey HM, Southwood S, Celis E, Wentworth PA, del Guercio MF, et al. Definition of an HLA-A3-like supermotif demonstrates the overlapping peptide-binding repertoires of common HLA molecules. *Hum Immunol*. (1996) 45:79–93. doi: 10.1016/0198-8859(95)00173-5
42. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. *BMC Immunol*. (2008) 9:1. doi: 10.1186/1471-2172-9-1
43. Sidney J, Southwood S, del Guercio MF, Grey HM, Chesnut RW, Kubo RT, et al. Specificity and degeneracy in peptide binding to HLA-B7-like class I molecules. *J Immunol*. (1996) 157:3480–90.
44. Sidney J, Southwood S, Mann DL, Fernandez-Vina MA, Newman MJ, Sette A. Majority of peptides binding HLA-A*0201 with high affinity crossreact with other A2-supertype molecules. *Hum Immunol*. (2001) 62:1200–16. doi: 10.1016/S0198-8859(01)00319-6
45. Sidney J, Southwood S, Paschetto V, Sette A. Simultaneous prediction of binding capacity for multiple molecules of the HLA B44 supertype. *J Immunol*. (2003) 171:5964–74. doi: 10.4049/jimmunol.171.11.5964
46. Sidney J, Southwood S, Sette A. Classification of A1- and A24-supertype molecules by analysis of their MHC-peptide binding repertoires. *Immunogenetics* (2005) 57:393–408. doi: 10.1007/s00251-005-0004-2

47. Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, et al. Several common HLA-DR types share largely overlapping peptide binding repertoires. *J Immunol.* (1998) 160:3363–73.
48. O'Sullivan D, Arrhenius T, Sidney J, Del Guercio MF, Albertson M, Wall M, et al. On the interaction of promiscuous antigenic peptides with different DR alleles. Identification of common structural motifs. *J Immunol.* (1991) 147:2663–9.
49. Castelli FA, Buhot C, Sanson A, Zarour H, Pouvelle-Moratille S, Nonn C, et al. HLA-DP4, the most frequent HLA II molecule, defines a new supertype of peptide-binding specificity. *J Immunol.* (2002) 169:6928–34. doi: 10.4049/jimmunol.169.12.6928
50. Berretta F, Butler RH, Diaz G, Sanarico N, Arroyo J, Fraziano M, et al. Detailed analysis of the effects of Glu/Lys beta69 human leukocyte antigen-DP polymorphism on peptide-binding specificity. *Tissue Antigens* (2003) 62:459–71. doi: 10.1046/j.1399-0039.2003.00131.x
51. Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different superotypes and a surprising degree of repertoire sharing across superotypes. *Immunogenetics* (2011) 63:325–35. doi: 10.1007/s00251-011-0513-0
52. Sidney J, Del Guercio MF, Southwood S, Sette A. The HLA Molecules DQA1*0501/B1*0201 and DQA1*0301/B1*0302 share an extensive overlap in peptide binding specificity. *J Immunol.* (2002) 169:5098–108. doi: 10.4049/jimmunol.169.9.5098
53. Tangri S, Mothe BR, Eisenbraun J, Sidney J, Southwood S, Briggs K, et al. Rationally engineered therapeutic proteins with reduced immunogenicity. *J Immunol.* (2005) 174:3187–96. doi: 10.4049/jimmunol.174.6.3187
54. Oseroff C, Sidney J, Kotturi MF, Kolla R, Alam R, Broide DH, et al. Molecular determinants of T cell epitope recognition to the common Timothy grass allergen. *J Immunol.* (2010) 185:943–55. doi: 10.4049/jimmunol.1000405
55. Oseroff C, Sidney J, Tripple V, Grey H, Wood R, Broide DH, et al. Analysis of T cell responses to the major allergens from German cockroach: epitope specificity and relationship to IgE production. *J Immunol.* (2012) 189:679–88. doi: 10.4049/jimmunol.1200694
56. Oseroff C, Sidney J, Vita R, Tripple V, McKinney DM, Southwood S, et al. T cell responses to known allergen proteins are differently polarized and account for a variable fraction of total response to allergen extracts. *J Immunol.* (2012) 189:1800–11. doi: 10.4049/jimmunol.1200850
57. Doolan DL, Southwood S, Chesnut R, Appella E, Gomez E, Richards A, et al. HLA-DR-promiscuous T cell epitopes from *Plasmodium falciparum* pre-erythrocytic-stage antigens restricted by multiple HLA class II alleles. *J Immunol.* (2000) 165:1123–37. doi: 10.4049/jimmunol.165.2.1123
58. Lamona V, Missale G, Urbani S, Pilli M, Boni C, Mori C, et al. Conserved hepatitis C virus sequences are highly immunogenic for CD4(+) T cells: implications for vaccine development. *Hepatology* (1999) 30:1088–98. doi: 10.1002/hep.510300435
59. Wilson CC, Palmer B, Southwood S, Sidney J, Higashimoto Y, Appella E, et al. Identification and antigenicity of broadly cross-reactive and conserved human immunodeficiency virus type 1-derived helper T-lymphocyte epitopes. *J Virol.* (2001) 75:4195–207. doi: 10.1128/JVI.75.9.4195-4207.2001
60. Arlehamn CS, Sidney J, Henderson R, Greenbaum JA, James EA, Moutafsi M, et al. Dissecting mechanisms of immunodominance to the common tuberculosis antigens ESAT-6, CFP10, Rv2031c (hspX), Rv2654c (TB7.7), and Rv1038c (EsxJ). *J Immunol.* (2012) 188:5020–31. doi: 10.4049/jimmunol.1103556
61. Assarsson E, Bui HH, Sidney J, Zhang Q, Glenn J, Oseroff C, et al. Immunomic analysis of the repertoire of T cell specificities for influenza A virus in humans. *J Virol.* (2008) 82:12241–51. doi: 10.1128/JVI.01563-08
62. Paul S, Lindestam Arlehamn CS, Scriba TJ, Dillon MB, Oseroff C, Hinz D, et al. Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. *J Immunol Methods* (2015) 422:28–34. doi: 10.1016/j.jim.2015.03.022
63. Dhanda SK, Karosiene E, Edwards L, Grifoni A, Paul S, Andreatta M, et al. Predicting HLA CD4 Immunogenicity in Human Populations. *Front Immunol.* (2018) 9:1369. doi: 10.3389/fimmu.2018.01369
64. Nielsen M, Lundegaard C, Lund O, Kesmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* (2005) 57:33–41. doi: 10.1007/s00251-005-0781-7
65. McKinney DM, Southwood S, Hinz D, Oseroff C, Arlehamn CS, Schulten V, et al. A strategy to determine HLA class II restriction broadly covering the DR, DP, and DQ allelic variants most commonly expressed in the general population. *Immunogenetics* (2013) 65:357–70. doi: 10.1007/s00251-013-0684-y
66. Paul S, Dillon MB, Arlehamn CS, Huang H, Davis MM, McKinney DM, et al. A population response analysis approach to assign class II HLA-epitope restrictions. *J Immunol.* (2015) 194:6164–76. doi: 10.4049/jimmunol.1403074
67. Paul S, Arlehamn CS, Schulten V, Westernberg L, Sidney J, Peters B, et al. Experimental validation of the RATE tool for inferring HLA restrictions of T cell epitopes. *BMC Immunol.* (2017) 18(Suppl. 1):20. doi: 10.1186/s12865-017-0204-1
68. John M, Gaudieri S. Influence of HIV and HCV on T cell antigen presentation and challenges in the development of vaccines. *Front Microbiol.* (2014) 5:514. doi: 10.3389/fmicb.2014.00514
69. Timm J, Walker CM. Mutational escape of CD8+ T cell epitopes: implications for prevention and therapy of persistent hepatitis virus infections. *Med Microbiol Immunol.* (2015) 204:29–38. doi: 10.1007/s00430-014-0372-z
70. Burdin N, Handy LK, Plotkin SA. What is wrong with pertussis vaccine immunity? the problem of waning effectiveness of pertussis vaccines. *Cold Spring Harb Perspect Biol.* (2017) 9:a029454. doi: 10.1101/cshperspect.a029454
71. Diavatopoulos DA, Edwards KM. What is wrong with pertussis vaccine immunity? why immunological memory to pertussis is failing. *Cold Spring Harb Perspect Biol.* (2017) 9:a029553. doi: 10.1101/cshperspect.a029553
72. Eberhardt CS, Siegrist CA. What is wrong with pertussis vaccine immunity? inducing and recalling vaccine-specific immunity. *Cold Spring Harb Perspect Biol.* (2017) 9:a029629. doi: 10.1101/cshperspect.a029629
73. Weiskopf D, Angelo MA, de Azeredo EL, Sidney J, Greenbaum JA, Fernando AN, et al. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proc Natl Acad Sci USA* (2013) 110:E2046–53. doi: 10.1073/pnas.1305227110
74. Weiskopf D, Angelo MA, Bangs DJ, Sidney J, Paul S, Peters B, et al. The human CD8+ T cell responses induced by a live attenuated tetravalent dengue vaccine are directed against highly conserved epitopes. *J Virol.* (2015) 89:120–8. doi: 10.1128/JVI.02129-14
75. Westernberg L, Schulten V, Greenbaum JA, Natali S, Tripple V, McKinney DM, et al. T-cell epitope conservation across allergen species is a major determinant of immunogenicity. *J Allergy Clin Immunol.* (2016) 138:571–8 e7. doi: 10.1016/j.jaci.2015.11.034
76. Pham J, Oseroff C, Hinz D, Sidney J, Paul S, Greenbaum J, et al. Sequence conservation predicts T cell reactivity against ragweed allergens. *Clin Exp Allergy* (2016) 46:1194–205. doi: 10.1111/cea.12772
77. Carrasco Pro S, Lindestam Arlehamn CS, Dhanda SK, Carpenter C, Lindvall M, Faruqi AA, et al. Microbiota epitope similarity either dampens or enhances the immunogenicity of disease-associated antigenic epitopes. *PLoS ONE* (2018) 13:e0196551. doi: 10.1371/journal.pone.0196551
78. Carrasco Pro S, Sidney J, Paul S, Lindestam Arlehamn CS, Weiskopf D, Peters B, et al. Automatic generation of validated specific epitope sets. *J Immunol Res.* (2015) 2015:763461. doi: 10.1155/2015/763461
79. Hinz D, Seumois G, Gholami AM, Greenbaum JA, Lane J, White B, et al. Lack of allergy to timothy grass pollen is not a passive phenomenon but associated with the allergen-specific modulation of immune reactivity. *Clin Exp Allergy* (2016) 46:705–19. doi: 10.1111/cea.12692
80. da Silva Antunes R, Pham J, McMurtrey C, Hildebrand WH, Phillips E, Mallal S, et al. Urinary peptides as a novel source of T cell allergen epitopes. *Front Immunol.* (2018) 9:886. doi: 10.3389/fimmu.2018.00886
81. da Silva Antunes R, Paul S, Sidney J, Weiskopf D, Dan JM, Phillips E, et al. Definition of human epitopes recognized in tetanus toxoid and development of an assay strategy to detect *ex vivo* tetanus CD4+ T cell responses. *PLoS ONE* (2017) 12:e0169086. doi: 10.1371/journal.pone.0169086
82. Weiskopf D, Angelo MA, Grifoni A, O'Rourke PH, Sidney J, Paul S, et al. HLA-DRB1 alleles are associated with different magnitudes of dengue

- virus-specific CD4+ T-Cell responses. *J Infect Dis.* (2016) 214:1117–24. doi: 10.1093/infdis/jiw309
83. Weiskopf D, Cerpas C, Angelo MA, Bangs DJ, Sidney J, Paul S, et al. Human CD8+ T cell responses against the four dengue virus serotypes are associated with distinct patterns of protein targets. *J Infect Dis.* (2015) 212:1743–51. doi: 10.1093/infdis/jiv289
 84. Grifoni A, Angelo MA, Lopez B, O'Rourke PH, Sidney J, Cerpas C, et al. Global assessment of dengue virus-specific CD4+ T cell responses in dengue-endemic areas. *Front Immunol.* (2017) 8:1309. doi: 10.3389/fimmu.2017.01309
 85. Hunt DF, Michel H, Dickinson TA, Shabanowitz J, Cox AL, Sakaguchi K, et al. Peptides presented to the immune system by the murine class II major histocompatibility complex molecule I-Ad. *Science* (1992) 256:1817–20. doi: 10.1126/science.1319610
 86. Cox AL, Skipper J, Chen Y, Henderson RA, Darrow TL, Shabanowitz J, et al. Identification of a peptide recognized by five melanoma-specific human cytotoxic T cell lines. *Science* (1994) 264:716–9. doi: 10.1126/science.7513441
 87. Caron E, Kowalewski DJ, Chiek Koh C, Sturm T, Schuster H, Aebersold R. Analysis of Major Histocompatibility Complex (MHC) immunopeptidomes using mass spectrometry. *Mol Cell Proteomics* (2015) 14:3105–17. doi: 10.1074/mcp.O115.052431
 88. Dan JM, Lindestam Arlehamn CS, Weiskopf D, da Silva Antunes R, Havenar-Daughton C, Reiss SM, et al. A cytokine-independent approach to identify antigen-specific human germinal center T follicular helper cells and rare antigen-specific CD4+ T cells in blood. *J Immunol.* (2016) 197:983–93. doi: 10.4049/jimmunol.1600318
 89. Bacher P, Schink C, Teutschbein J, Knienmeyer O, Assenmacher M, Brakhage AA, et al. Antigen-reactive T cell enrichment for direct, high-resolution analysis of the human naive and memory Th cell repertoire. *J Immunol.* (2013) 190:3967–76. doi: 10.4049/jimmunol.1202221
 90. Bacher P, Scheffold A. Flow-cytometric analysis of rare antigen-specific T cells. *Cytometry Part A* (2013) 83:692–701. doi: 10.1002/cyto.a.22317
 91. Mallone R, Nepom GT. MHC Class II tetramers and the pursuit of antigen-specific T cells: define, deviate, delete. *Clin Immunol.* (2004) 110:232–42. doi: 10.1016/j.clim.2003.11.004
 92. Nepom GT. MHC multimers: expanding the clinical toolkit. *Clin Immunol.* (2003) 106:1–4. doi: 10.1016/S1521-6616(02)00014-1
 93. Kotturi MF, Swann JA, Peters B, Arlehamn CL, Sidney J, Kolla RV, et al. Human CD8(+) and CD4(+) T cell memory to lymphocytic choriomeningitis virus infection. *J Virol.* (2011) 85:11770–80. doi: 10.1128/JVI.05477-11
 94. Burel JG, Qian Y, Lindestam Arlehamn C, Weiskopf D, Zapardiel-Gonzalo J, Taplitz R, et al. An integrated workflow to assess technical and biological variability of cell population frequencies in human peripheral blood by flow cytometry. *J Immunol.* (2017) 198:1748–58. doi: 10.4049/jimmunol.1601750
 95. Tian Y, Babor M, Lane J, Schulten V, Patil VS, Seumois G, et al. Unique phenotypes and clonal expansions of human CD4 effector memory T cells re-expressing CD45RA. *Nat Commun.* (2017) 8:1473. doi: 10.1038/s41467-017-01728-5
 96. Weiskopf D, Bangs DJ, Sidney J, Kolla RV, De Silva AD, de Silva AM, et al. Dengue virus infection elicits highly polarized CX3CR1+ cytotoxic CD4+ T cells associated with protective immunity. *Proc Natl Acad Sci USA.* (2015) 112:E4256–63. doi: 10.1073/pnas.1505956112
 97. Spitzer MH, Nolan GP. Mass Cytometry: Single Cells, Many Features. *Cell* (2016) 165:780–91. doi: 10.1016/j.cell.2016.04.019
 98. Diggins KE, Ferrell PB Jr, Irish JM. Methods for discovery and characterization of cell subsets in high dimensional mass cytometry data. *Methods* (2015) 82:55–63. doi: 10.1016/j.ymeth.2015.05.008
 99. Golumbeanu M, Cristinelli S, Rato S, Munoz M, Cavassini M, Beerenwinkel N, et al. Single-cell RNA-Seq reveals transcriptional heterogeneity in latent and reactivated HIV-infected cells. *Cell Rep.* (2018) 23:942–50. doi: 10.1016/j.celrep.2018.03.102
 100. Kashima Y, Suzuki A, Liu Y, Hosokawa M, Matsunaga H, Shirai M, et al. Combinatory use of distinct single-cell RNA-seq analytical platforms reveals the heterogeneous transcriptome response. *Sci Rep.* (2018) 8:3482. doi: 10.1038/s41598-018-21161-y
 101. Russell AB, Trapnell C, Bloom JD. Extreme heterogeneity of influenza virus infection in single cells. *Elife* (2018) 7:e32303. doi: 10.7554/eLife.32303
 102. Patil VS, Madrigal A, Schmiedel BJ, Clarke J, O'Rourke P, de Silva AD, et al. Precursors of human CD4(+) cytotoxic T lymphocytes identified by single-cell transcriptome analysis. *Sci Immunol* (2018) 3:eaan8664. doi: 10.1126/sciimmunol.aan8664
 103. Tian Y, Sette A, Weiskopf D. Cytotoxic CD4 T cells: differentiation, function, and application to dengue virus infection. *Front Immunol.* (2016) 7:531. doi: 10.3389/fimmu.2016.00531
 104. Rosati E, Dowds CM, Liaskou E, Henriksen EKK, Karlens TH, Franke A. Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnol.* (2017) 17:61. doi: 10.1186/s12896-017-0379-9
 105. Glanville J, Huang H, Nau A, Hatton O, Wagar LE, Rubelt F, et al. Identifying specificity groups in the T cell receptor repertoire. *Nature* (2017) 547:94–8. doi: 10.1038/nature22976
 106. Dash P, Fiore-Gartland AJ, Hertz T, Wang GC, Sharma S, Souquette A, et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* (2017) 547:89–93. doi: 10.1038/nature22383
 107. Mills KH. Immunity to *Bordetella pertussis*. *Microbes and infection / Institut Pasteur* (2001) 3:655–77. doi: 10.1016/S1286-4579(01)01421-6
 108. Mooi FR, NA VDM, De Melker HE. Pertussis resurgence: waning immunity and pathogen adaptation - two sides of the same coin. *Epidemiol Infect.* (2014) 142:685–94. doi: 10.1017/S0950268813000071
 109. Fulton TR, Phadke VK, Orenstein WA, Hinman AR, Johnson WD, Omer SB. Protective effect of contemporary pertussis vaccines: a systematic review and meta-analysis. *Clin Infect Dis.* (2016) 62:1100–10. doi: 10.1093/cid/ciw051
 110. Sheridan SL, Frith K, Snelling TL, Grimwood K, McIntyre PB, Lambert SB. Waning vaccine immunity in teenagers primed with whole cell and acellular pertussis vaccine: recent epidemiology. *Expert Rev Vaccines* (2014) 13:1081–106. doi: 10.1586/14760584.2014.944167
 111. Gu XX, Plotkin SA, Edwards KM, Sette A, Mills KHG, Levy O, et al. Waning immunity and microbial vaccines-workshop of the national institute of allergy and infectious diseases. *Clin Vaccine Immunol* (2017) 24:e00034–17. doi: 10.1128/CI.00034-17
 112. Vaughan K, Seymour E, Peters B, Sette A. Substantial gaps in knowledge of *Bordetella pertussis* antibody and T cell epitopes relevant for natural immunity and vaccine efficacy. *Hum Immunol.* (2014) 75:440–51. doi: 10.1016/j.humimm.2014.02.013
 113. Barkoff AM, Grondahl-Yli-Hannuksela K, Vuonoirta J, Mertsola J, Kallonen T, He Q. Differences in avidity of IgG antibodies to pertussis toxin after acellular pertussis booster vaccination and natural infection. *Vaccine* (2012) 30:6897–902. doi: 10.1016/j.vaccine.2012.09.003
 114. Murphy BR, Walsh EE. Formalin-inactivated respiratory syncytial virus vaccine induces antibodies to the fusion glycoprotein that are deficient in fusion-inhibiting activity. *J Clin Microbiol.* (1988) 26:1595–7.
 115. Bart MJ, van Gent M, van der Heide HG, Boekhorst J, Hermans P, Parkhill J, et al. Comparative genomics of prevaccination and modern *Bordetella pertussis* strains. *BMC Genomics* (2010) 11:627. doi: 10.1186/1471-2164-11-627
 116. Octavia S, Sintchenko V, Gilbert GL, Lawrence A, Keil AD, Hogg G, et al. Newly emerging clones of *Bordetella pertussis* carrying prn2 and ptxP3 alleles implicated in Australian pertussis epidemic in 2008–2010. *J Infect Dis.* (2012) 205:1220–4. doi: 10.1093/infdis/jis178
 117. Octavia S, Maharjan RP, Sintchenko V, Stevenson G, Reeves PR, Gilbert GL, et al. Insight into evolution of *Bordetella pertussis* from comparative genomic analysis: evidence of vaccine-driven selection. *Mol Biol Evol.* (2011) 28:707–15. doi: 10.1093/molbev/msq245
 118. Kurniawan J, Maharjan RP, Chan WF, Reeves PR, Sintchenko V, Gilbert GL, et al. *Bordetella pertussis* clones identified by multilocus variable-number tandem-repeat analysis. *Emerg Infect Dis.* (2010) 16:297–300. doi: 10.3201/eid1602.081707
 119. Mooi FR, van Oirschot H, Heuvelman K, van der Heide HG, Gaastra W, Willems RJ. Polymorphism in the *Bordetella pertussis* virulence factors P.69/pertactin and pertussis toxin in The Netherlands: temporal trends and evidence for vaccine-driven evolution. *Infect Immun.* (1998) 66:670–5.
 120. Le T, Cherry JD, Chang SJ, Knoll MD, Lee ML, Barenkamp S, et al. Immune responses and antibody decay after immunization of adolescents and adults with an acellular pertussis vaccine: the APERT Study. *J Infect Dis.* (2004) 190:535–44. doi: 10.1086/422035

121. Dalby T, Petersen JW, Harboe ZB, Krogfelt KA. Antibody responses to pertussis toxin display different kinetics after clinical *Bordetella pertussis* infection than after vaccination with an acellular pertussis vaccine. *J Med Microbiol.* (2010) 59(Pt 9):1029–36. doi: 10.1099/jmm.0.020826-0
122. Heininger U, Cherry JD, Stehr K. Serologic response and antibody-titer decay in adults with pertussis. *Clin Infect Dis.* (2004) 38:591–4. doi: 10.1086/381439
123. Barnard A, Mahon BP, Watkins J, Redhead K, Mills KH. Th1/Th2 cell dichotomy in acquired immunity to *Bordetella pertussis*: variables in the *in vivo* priming and *in vitro* cytokine detection techniques affect the classification of T-cell subsets as Th1, Th2 or Th0. *Immunology* (1996) 87:372–80. doi: 10.1046/j.1365-2567.1996.497560.x
124. Warfel JM, Merkel TJ. *Bordetella pertussis* infection induces a mucosal IL-17 response and long-lived Th17 and Th1 immune memory cells in nonhuman primates. *Mucosal Immunol.* (2013) 6:787–96. doi: 10.1038/mi.2012.117
125. Warfel JM, Zimmerman LI, Merkel TJ. Acellular pertussis vaccines protect against disease but fail to prevent infection and transmission in a nonhuman primate model. *Proc Natl Acad Sci USA.* (2014) 111:787–92. doi: 10.1073/pnas.1314688110
126. Rowe J, Macaubas C, Monger TM, Holt BJ, Harvey J, Poolman JT, et al. Antigen-specific responses to diphtheria-tetanus-acellular pertussis vaccine in human infants are initially Th2 polarized. *Infect Immun.* (2000) 68:3873–7. doi: 10.1128/IAI.68.7.3873-3877.2000
127. Rowe J, Yerkovich ST, Richmond P, Suriyaarachchi D, Fisher E, Feddema L, et al. Th2-associated local reactions to the acellular diphtheria-tetanus-pertussis vaccine in 4- to 6-year-old children. *Infect Immun.* (2005) 73:8130–5. doi: 10.1128/IAI.73.12.8130-8135.2005
128. Ryan EJ, Nilsson L, Kjellman N, Gothefors L, Mills KH. Booster immunization of children with an acellular pertussis vaccine enhances Th2 cytokine production and serum IgE responses against pertussis toxin but not against common allergens. *Clin Exp Immunol.* (2000) 121:193–200. doi: 10.1046/j.1365-2249.2000.01306.x
129. Ausiello CM, Urbani F, la Sala A, Lande R, Cassone A. Vaccine- and antigen-dependent type 1 and type 2 cytokine induction after primary vaccination of infants with whole-cell or acellular pertussis vaccines. *Infect Immun.* (1997) 65:2168–74.
130. Chiappini E, Stival A, Galli L, de Martino M. Pertussis re-emergence in the post-vaccination era. *BMC Infect Dis.* (2013) 13:151. doi: 10.1186/1471-2334-13-151
131. Edwards KM, Berbers GA. Immune responses to pertussis vaccines and disease. *J Infect Dis.* (2014) 209 (Suppl. 1):S10–5. doi: 10.1093/infdis/jit560
132. Plotkin SA. The pertussis problem. *Clin Infect Dis.* (2014) 58:830–3. doi: 10.1093/cid/cit934
133. van der Lee S, Hendriks LH, Sanders EAM, Berbers GAM, Buisman AM. Whole-cell or acellular pertussis primary immunizations in infancy determines adolescent cellular immune profiles. *Front Immunol.* (2018) 9:51. doi: 10.3389/fimmu.2018.00051
134. Carbonetti NH, Artamonova GV, Andreassen C, Bushar N. Pertussis toxin and adenylate cyclase toxin provide a one-two punch for establishment of *Bordetella pertussis* infection of the respiratory tract. *Infect Immun.* (2005) 73:2698–703. doi: 10.1128/IAI.73.5.2698-2703.2005
135. Carbonetti NH. Pertussis toxin and adenylate cyclase toxin: key virulence factors of *Bordetella pertussis* and cell biology tools. *Future Microbiol.* (2010) 5:455–69. doi: 10.2217/fmb.09.133
136. Boehm DT, Hall JM, Wong TY, DiVenere A, Sen-Kilic E, Bevere JR, et al. Evaluation of adenylate cyclase toxoid antigen in acellular pertussis vaccines using a *Bordetella pertussis* challenge model in mice. *Infect Immun.* (2018) 86:e00857–17. doi: 10.1128/IAI.00857-17
137. Warfel JM, Edwards KM. Pertussis vaccines and the challenge of inducing durable immunity. *Curr Opin Immunol.* (2015) 35:48–54. doi: 10.1016/j.coi.2015.05.008
138. Warfel JM, Zimmerman LI, Merkel TJ. Comparison of three whole-cell pertussis vaccines in the baboon model of pertussis. *Clin Vaccine Immunol.* (2015) 23:47–54. doi: 10.1128/CVI.00449-15
139. Hamid R, Brandt SJ. Transforming growth-interacting factor (TGIF) regulates proliferation and differentiation of human myeloid leukemia cells. *Mol Oncol.* (2009) 3:451–63. doi: 10.1016/j.molonc.2009.07.004
140. Melhuish TA, Gallo CM, Wotton D. TGIF2 interacts with histone deacetylase 1 and represses transcription. *J Biol Chem.* (2001) 276:32109–14. doi: 10.1074/jbc.M103377200
141. Bresciani A, Paul S, Schommer N, Dillon MB, Bancroft T, Greenbaum J, et al. T-cell recognition is shaped by epitope sequence conservation in the host proteome and microbiome. *Immunology* (2016) 148:34–9. doi: 10.1111/imm.12585
142. Liepe J, Marino F, Sidney J, Jeko A, Bunting DE, Sette A, et al. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science* (2016) 354:354–8. doi: 10.1126/science.aaf4384
143. Faridi P, Li C, Ramarathnam SH, Vivian JP, Illing PT, Mifsud NA, et al. A subset of HLA-I peptides are not genomically templated: evidence for cis- and trans-spliced peptide ligands. *Sci Immunol.* (2018) 3:eaar3947. doi: 10.1126/sciimmunol.aar3947

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Tian, da Silva Antunes, Sidney, Lindestam Arlehamn, Grifoni, Dhanda, Paul, Peters, Weiskopf and Sette. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Trained Immunity-Based Vaccines: A New Paradigm for the Development of Broad-Spectrum Anti-infectious Formulations

Silvia Sánchez-Ramón^{1,2*}, Laura Conejero³, Mihai G. Netea^{4,5}, David Sancho⁶, Óscar Palomares⁷ and José Luis Subiza³

¹ Department of Clinical Immunology and IdISSC, Hospital Clínico San Carlos, Madrid, Spain, ² Department of Immunology, ENT and Ophthalmology, Complutense University School of Medicine, Madrid, Spain, ³ Immunotek, Alcalá de Henares, Spain, ⁴ Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, Netherlands, ⁵ Department for Genomics and Immunoregulation, Life and Medical Sciences Institute, University of Bonn, Bonn, Germany, ⁶ Immunobiology Laboratory, Centro Nacional de Investigaciones Cardiovasculares, Madrid, Spain, ⁷ Department of Biochemistry and Molecular Biology, School of Chemistry, Complutense University of Madrid, Madrid, Spain

OPEN ACCESS

Edited by:

Pedro A. Reche,
Complutense University of Madrid,
Spain

Reviewed by:

Randy A. Albrecht,
Icahn School of Medicine at Mount
Sinai, United States
Michael Schotsaert,
Icahn School of Medicine at Mount
Sinai, United States

*Correspondence:

Silvia Sánchez-Ramón
ssramon@salud.madrid.org

Specialty section:

This article was submitted to
Vaccines and Molecular Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 04 September 2018

Accepted: 29 November 2018

Published: 17 December 2018

Citation:

Sánchez-Ramón S, Conejero L,
Netea MG, Sancho D, Palomares Ó
and Subiza JL (2018) Trained
Immunity-Based Vaccines: A New
Paradigm for the Development of
Broad-Spectrum Anti-infectious
Formulations. *Front. Immunol.* 9:2936.
doi: 10.3389/fimmu.2018.02936

Challenge with specific microbial stimuli induces long lasting epigenetic changes in innate immune cells that result in their enhanced response to a second challenge by the same or unrelated microbial insult, a process referred to as trained immunity. This opens a new avenue in vaccinology to develop *Trained Immunity-based Vaccines* (TlbV), defined as vaccine formulations that induce training in innate immune cells. Unlike conventional vaccines, which are aimed to elicit only specific responses to vaccine-related antigens, TlbV aim to stimulate broader responses. As trained immunity is generally triggered by pattern recognition receptors (PRRs), TlbV should be formulated with microbial structures containing suitable PRR-ligands. The TlbV concept we describe here may be used for the development of vaccines focused to promote host resistance against a wide spectrum of pathogens. Under the umbrella of trained immunity, a broad protection can be achieved by: (i) increasing the nonspecific effector response of innate immune cells (e.g., monocyte/macrophages) to pathogens, (ii) harnessing the activation state of dendritic cells to enhance adaptive T cell responses to both specific and nonrelated (bystander) antigens. This capacity of TlbV to promote responses beyond their nominal antigens may be particularly useful when conventional vaccines are not available or when multiple coinfections and/or recurrent infections arise in susceptible individuals. As the set of PRR-ligands chosen is essential not only for stimulating trained immunity but also to drive adaptive immunity, the precise design of TlbV will improve with the knowledge on the functional relationship among the different PRRs. While the TlbV concept is emerging, a number of the current anti-infectious vaccines, immunostimulants, and even vaccine adjuvants may already fall in the TlbV category. This may apply to increase immunogenicity of novel vaccine design approaches based on small molecules, like those achieved by reverse vaccinology.

Keywords: adjuvants, innate immunity, immunostimulants, pattern recognition receptors (PRRs), PRR-ligands, trained immunity, trained immunity-based vaccines (TlbV), vaccines

BACKGROUND

Conventional anti-infectious vaccines are primarily intended to target specific pathogens by enhancing an antigen-specific adaptive immune response. This response is based on triggering B and T lymphocytes that, by virtue of their clonally segregated antigen receptors, generate effector, and memory cells. Proliferation and differentiation of specific lymphocytes is the basis of immunological memory, a hallmark of the adaptive immune response and the rationale behind conventional vaccines. Louis Pasteur built on the work of Edward Jenner to develop the principles of using attenuated live microbes to prevent the pathogen's caused disease (1). Pre-exposure vaccination was a major breakthrough in the prevention of many infectious diseases. His rabies vaccine for post-exposure prophylaxis in the severely ill boy Joseph Meister in 1885 raised the concept of therapeutic vaccines (2).

Immunological adaptive memory has been traditionally defined as long-term acquired memory against an encountered antigen through infection or immunization, leading to a quicker and heightened immune response upon an ulterior *rendezvous* (3). Resting clones of memory B and T cells can survive at different compartments for several decades until reactivation by recall responses (4, 5). Recent epidemiological studies highlight the role of subclinical infections or repeated endemic exposure for the maintenance of protective antigen-specific antibodies and T cells, indicating the dependency of this adaptive memory on antigen-re-exposure (6, 7). Besides, the persistent specific T and B lymphocyte activation can also favor “infectious immunity,” a process by which innate immune responses are enhanced by mechanisms depending on the persistence of the activation of adaptive immunity (3).

Trained Immunity

Importantly however, solid epidemiological data have also demonstrated that certain mild infections or vaccinations, such as with bacilli Calmette-Guerin (BCG), lead to protection against heterologous infections, with a strong impact on overall mortality due to infection for up to 1 year (8–11). When vaccination against smallpox was introduced around 200 years ago, positive side-effects such as protection against measles, scarlet fever and whooping cough, among others, were noticed (12). These and many other clinical observations, pointed to a long-lasting non-specific collateral benefit associated to these vaccines, regardless of specific priming and subsequent clonal selection of T and B lymphocytes specific for the nominal antigens present in the vaccine. In recent years, it has become evident that cells of the innate immunity may be primed upon encounter with certain pathogens or molecular patterns associated to pathogens (PAMPs), acquiring a higher resistance to a second infection against the same or unrelated pathogens (cross-protection) for a relatively long time (13, 14). This concept has been termed “trained (innate) immunity.” Trained immunity implies adaptation of innate immunity processes in a *de-facto* innate immunological memory, and plays an essential role in vertebrates (15), which is similar to that described for bacteria, plants and invertebrates (16).

Mechanistically, trained immunity is defined by immunological, metabolic and epigenetic hallmarks (17–20). Several studies have shown that metabolic reprogramming through a shift from oxidative phosphorylation to aerobic glycolysis (the Warburg effect) mediated by the Akt/mTOR/HIF-1 α pathway is a key mechanism for trained immunity responses (21, 22). The glycolysis, glutaminolysis, and cholesterol synthesis pathways in monocytes and macrophages were identified as the essential underlying mechanism linking epigenetic rewiring and the induction of improved innate immunity (22, 23). Thus, changes in cellular metabolism influence the epigenetic reprogramming of innate immune cells having an impact on cytokine and reactive oxygen species' production. In this regard, trained immunity regulates epigenetic changes such as H3K4 trimethylation and H3K27 acetylation, both associated with active chromatin, and H3K9 trimethylation, a repressive marker (19, 22, 24). Trained immunity favors the production and release of proinflammatory cytokines such as TNF- α , IL-6 and IL-1 β by innate immune cells upon exposure to a second stimulus (14, 24, 25). Most of these features differ from what classically has been postulated for the innate immune system, as trained immunity induces functional reprogramming within innate immune cells that maintain these cells in a “ready-to-react” functional state over extended periods of time. Interestingly, although a maximum duration of trained immunity effects has been reported up to 3 months (24), a long lasting effect of trained cells with the capacity to enhance T cell responses up to 1 year is feasible (26), thus bridging innate training with adaptive responses. Moreover, the storage of specific long peptides for ulterior long-lasting cross-presentation to elicit cytotoxic T lymphocytes is an additional feature that may be associated to trained immunity in monocytes and could bridge innate and adaptive imprinting (27).

A striking example indicating a durable change within the innate immunity compartment is the imprinting of BCG on bone marrow hematopoietic stem cells and multipotent progenitors, giving rise to epigenetically modified macrophages that provide better protection against virulent *M. tuberculosis* than naïve macrophages (28). Unlike the classical memory following the adaptive immune response, long-term responses associated to trained immunity are not based on a clonal expansion of lymphocytes but on reprogramming myeloid cells by stable epigenetic changes (Table 1).

Pathogen recognition receptors (PRRs) expressed on innate immune cells, including long-lived macrophages and their precursors, are involved in the stimulation of trained immunity. Different PRRs have been involved in this task, such as C-type lectin receptors (CLRs) and Nod-like receptors (NLRs). Training of the innate immunity is therefore based on boosting non-specific immunity to re-infection by bacteria, fungi or viruses by certain pathogen's derived components (19). There are many examples of pathogen-associated molecules with evidence of cross-protection in experimental models (Table 2). The increased host defense induced by trained immunity, while effective against a range of pathogens, is non-specific as it is mediated by the release of proinflammatory cytokines such as IL-1 α and TNF- α and/or reactive oxygen species (ROS) (17). The

TABLE 1 | Similarities and differences between trained and adaptive immunity.

Feature	Trained immunity	Adaptive immunity
Specificity	±	+++
Inducers	Pathogens and derived products (PAMPs)	Antigen presentation plus costimulatory signals and cytokines from DCs
Receptors	PRRs	slg, TCR, receptors for costimulation and cytokines
Clonality	No	Yes
Cells	Monocytes, Macrophages, NK, DCs, ILCs, and other innate immune cells	B and T lymphocytes
Memory	Months	Years
Memory mechanism	Epigenetic modifications	Clonal expansion and differentiation

TABLE 2 | Examples of pathogen-associated molecules with experimental evidence of cross-protection.

Component	Source	Cross-protection	References
LPS (endotoxin)	Most Gram-negative bacteria, such as <i>E. coli</i>	<i>Staphylococcus aureus</i>	(29)
Peptidoglycan component muramyl dipeptide	Bacteria	Toxoplasma	(30)
Flagellin	Gram-negative bacteria	Gram-positive bacterium <i>Streptococcus pneumoniae</i>	(31)
FimH	<i>E. coli</i>	Rotavirus	(32)
β-glucan	Fungi	Influenza virus	(33)
		<i>Staphylococcus aureus</i>	(34)
		<i>Streptococcus pneumoniae</i>	
Chitin	Fungi	<i>Staphylococcus aureus</i> or <i>Escherichia coli</i>	(35)
CpG oligodeoxynucleotide	Bacteria, synthetic	<i>E. coli</i>	(36)
		Influenza virus	(37)
			(38)

role of IL-1 α in host resistance to infection and how TNF- α protects against infections have been recently updated (39, 40). Higher resistance due to trained immunity does not mean however an absolute resistance to every type of second infection that, on the other hand, may be favored by facts beyond the innate immune response. This might account for why a natural infection, such as primary infection by influenza A virus, can result in bacterial pneumonia resulting from superinfection by *Streptococcus pneumoniae* or other bacterial strains.

MOVING AHEAD CONVENTIONAL VACCINES: TRAINED IMMUNITY-BASED VACCINES (TIBV)

The exploitation of the principles of trained immunity may result in a next generation of anti-infectious vaccines (41–43). Trained immunity-based vaccines (TibV) may confer a broad protection far beyond to the nominal antigens they contain. By proper targeting of innate immune cells to stimulate trained immunity, both nonspecific and specific immune responses can be enhanced by TibV. Such responses can also be driven against bystander pathogens encountered by the host during the window of trained immunity.

Vaccines using attenuated and/or inactivated pathogens may be examples of TibV as long as they contain PAMPs able to trigger PRRs inducing trained immunity. Different PRR ligands have been described as trained immunity stimuli, like *Candida*-derived β -glucan or BCG-derived muramyl dipeptide, triggering CLRs (dectin-1) or NLRs (NOD2), respectively. Of note, by using these training stimuli, slight differences on how they modify the cellular metabolism of innate immune cells has been described (22). This opens the possibility that different trained immunity outcomes may be achieved varying the set of PRR ligands used in the TibV. At this point, it should be noted that, within the context of a vaccine, the fact that trained innate immune cells may enhance adaptive responses is essential. In this regard, the role played by DCs, the cellular link between innate and adaptive immunity (44), must be pivotal. Although trained immunity is linked to innate cells such as monocytes, macrophages and NK cells (19), the trained immunity-promoting BCG vaccine can also enhance heterologous T cell responses (26, 45, 46). It has been speculated that the increased expression of certain PRR in innate trained cells, as well as the release of typical innate immunity cytokines, such as IL-1 β , contribute to enhance adaptive T cell responses (26). In this regard, the polybacterial sublingual vaccine MV130, which contains whole cell heat-inactivated bacteria, has been shown to enhance *in vivo* T cell responses to unrelated antigens, while priming DCs and inducing IL-1 β release *in vitro* (47). DCs with high immunostimulatory properties that enhance adaptive immune responses via IL-1 β release have been described (48). Moreover, the role of inflammasome-associated IL-1 family cytokines in delineating the adaptive immune responses is established regarding differentiation of Th17 cells and promoting effector functions of Th1 cells (49). Interestingly, these DCs are rendered “hyperactive” by releasing IL-1 β in absence of cell death by virtue of an alternative inflammasome pathway modulated by certain TLR ligands of microbial origin, like LPS and peptidoglycans (49). Thus, strong adaptive immune responses may be stimulated by certain PRRs stimuli that may complement those inducing trained immunity.

One of the most interesting aspects of trained immunity is that innate immune cells maintain a primed functional state for a quite long period of time (50). As it may last more than several months (26), the enhanced immune responses induced by a given TibV may also be applied to possible bystander pathogens encountered by the host during this time frame. Thus,

another relevant aspect of TIbV is that they may promote both nonspecific and specific resistance to unrelated pathogens while trained immunity is still present. This may be of particular interest when co-infections with the pathogens included in the TIbV are likely, especially in the context of recurrent infections. Thus, under the umbrella of trained immunity, the TIbV concept emerges as a new paradigm of vaccines seeking to increase host resistance against a broad spectrum of pathogens.

With the above concepts in mind, the proposed term of TIbV can be applied for those anti-infectious vaccines composed of whole microorganisms or derived products that display the following features (**Figure 1**):

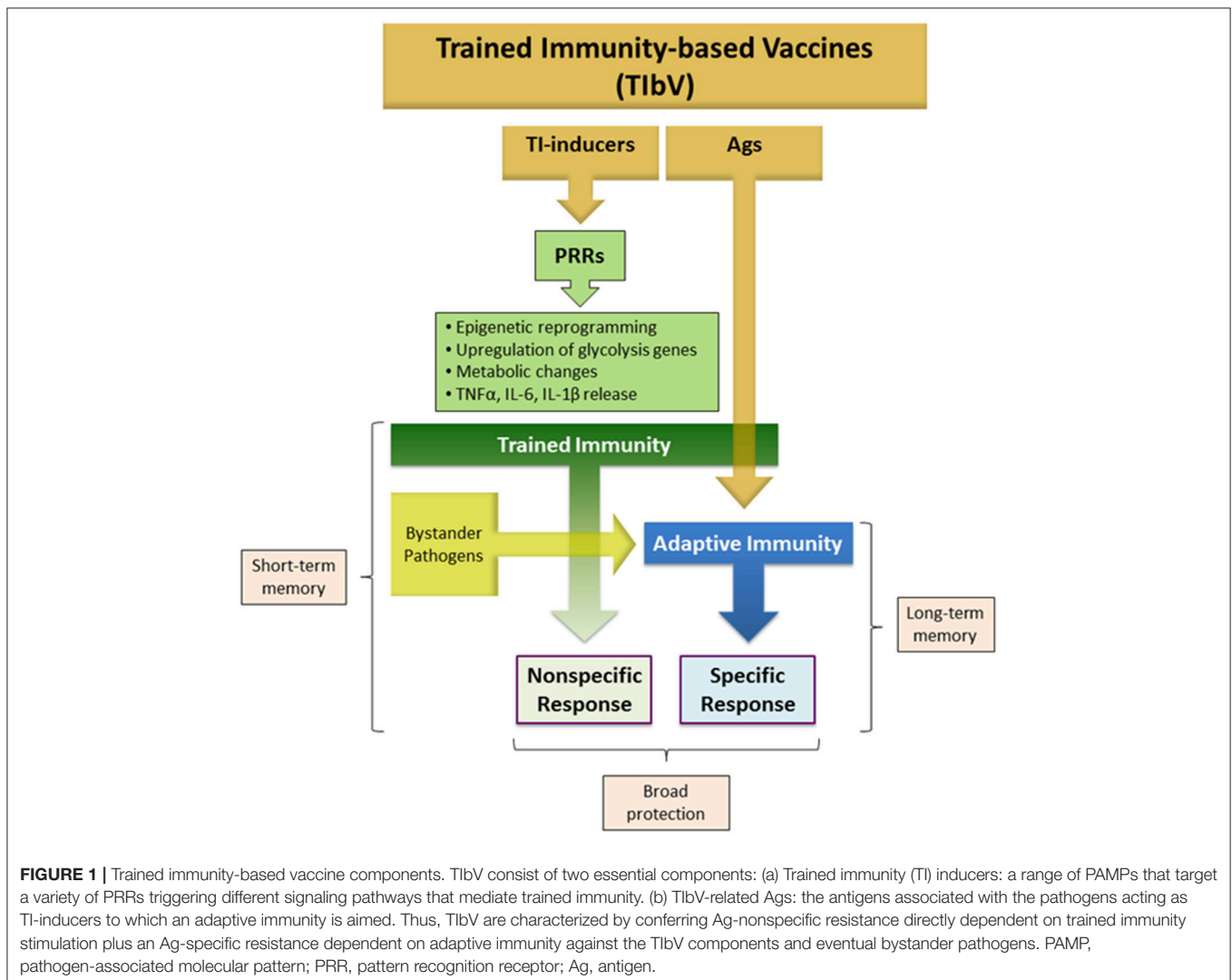
- i) Composed of trained immunity inducers, i.e. certain PAMPs, in addition to pathogen specific antigens.
- ii) Intended to be effective not only against the specific pathogens targeted by the vaccine, but also against heterologous pathogens.

In contrast to conventional vaccines, TIbV efficacy cannot be measured solely in terms of specific responses to the nominal

antigens included in the vaccine. In this regard, the clinical outcome scored by lower infectious rates in particular clinical settings is necessary.

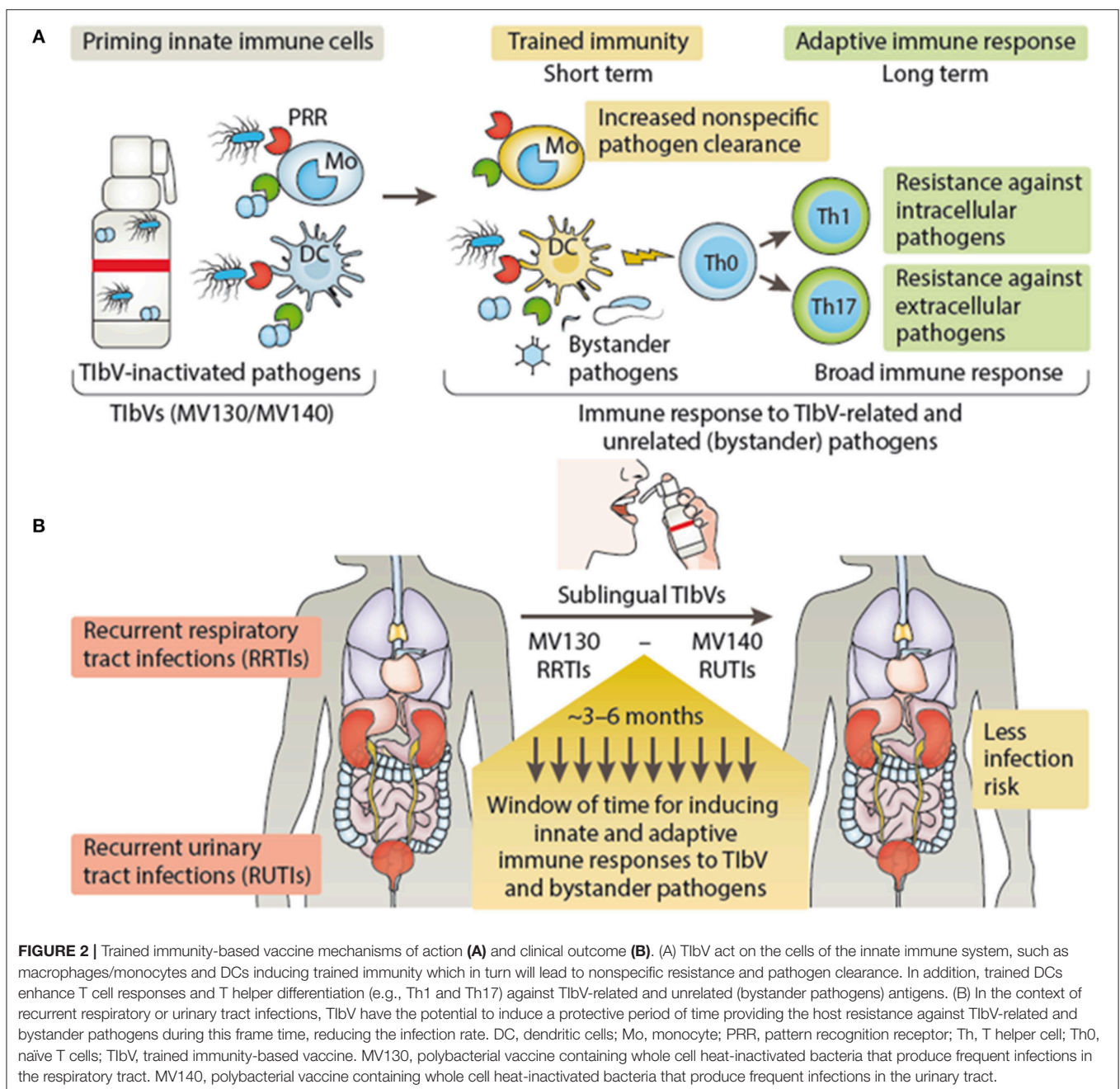
Examples of Non-conventional Vaccines That Can be Ascribed as TIbV

Vaccines that may fall in the category of TIbV include those bacterial preparations used for recurrent infections for either the respiratory or urinary tract (51–53). Previous and recent studies provide many clinical observations that combinations of inactivated bacterial vaccines induce cross-protection against infections produced by quite different microorganisms (41, 54). In the case of the sublingual vaccine MV130, designed to prevent recurrent respiratory tract infections, a significant reduction in patient's rate of infection was observed (51). Besides inducing specific T cell immunity against the bacteria included in MV130, treated patients showed an enhancement in T cell response to unrelated flu antigens (51). MV130 triggers TLR and NLR signaling pathways on DCs releasing trained immunity hallmark cytokines (TNF α , IL-6, and IL-1 β) (47). In addition, MV130



promoted the generation of Th1 and Th17 responses with high levels of IL-10 both *in vitro* and *in vivo* to MV130-related and bystander antigens (47). A recent clinical trial performed in children with recurrent wheezing attacks (mostly of viral etiology) has shown the clinical benefit of MV130 as well as the protection in experimental models of respiratory viral infections by trained immunity mechanisms (Nieto et al., manuscript in preparation). Similarly, MV140 another sublingual whole cell heat-inactivated bacterial vaccine designed to prevent recurrent urinary tract infections (52, 53, 55), was also effective against urobacteria species not included in its composition (53). MV140

also triggers the release of TNF α , IL-6 and IL-1 β by DCs, albeit using different signaling pathways from MV130, and induces Th1 and Th17 responses by mechanisms mediated by CLRs and TLRs as well (56). Thus, both MV130 and M140 vaccines induce the release of a similar set of cytokines ascribed to trained immunity, and favor heterologous Th1 and Th17 responses *in vivo* as described for vaccines stimulating trained immunity (26). **Figure 2** summarizes the proposed action mechanisms of either MV130 or MV140 as putative TibV. Other bacterial preparations used as nonspecific immunostimulants may also act as trained immunity inducers (see below).



Conventional Vaccines With Associated Trained-Immunity Effects as Potential TIbV

A number of conventional anti-infectious vaccines, most of them containing live-attenuated pathogens, have been shown to induce, in addition to the intended specific memory, broad protection by nonspecific mechanisms (46, 58). Therefore, they can be considered within the category of TIbV provided that such mechanisms are related to trained immunity. If this were the case, harnessing innate memory as part of a vaccination strategy with these vaccines may be considered.

BCG

As above mentioned, most studies that examined and elucidated the mechanisms of trained immunity have been performed with BCG as a model. Randomized-controlled trials carried out in Guinea-Bissau with low-birth-weight infants early vaccinated against tuberculosis with BCG, demonstrated clear beneficial effects reducing all-cause mortality, especially due to neonatal sepsis, respiratory infections, and fever (9). The effects of neonatal BCG vaccination on T and B lymphocytes subsets in infants in Denmark showed limited impact though (57) and did not affect parent-reported infections (59). It is not known whether the clinical setting in different populations, exposed to a high vs. low rate of pathogens, might account for these divergent outcomes. Recently, Arts et al. have demonstrated that BCG vaccination confers protection against viral infection (25). In a placebo-controlled clinical trial with BCG, all volunteers received the yellow fever vaccine 1 month after BCG, as an experimental mild viral infection. BCG-vaccinated volunteers displayed a significant reduction of viremia compared to the placebo group, which highly correlated with enhanced IL-1 β production (25). In some experimental models of virus infection BCG immunization has been reported to confer non-specific protection; yet not in all, pointing out that the route and dose of BCG administration may be important (60).

BCG is currently used as local immunotherapy in bladder cancer (61). Interestingly, Buffen and cols. have demonstrated that the anticancer effects of BCG were dependent on trained immunity (62). In addition to a nonspecific cytotoxic effect for tumor cells by BCG-trained innate cells, these may enhance tumor specific T cell responses as a massive accumulation of tumor specific T lymphocytes are recovered in urine after successful BCG therapy (63). As endogenous tumor antigens may act as bystander antigens under the umbrella of trained immunity, this opens the possibility of TIbV as immunostimulants outside of anti-infectious purposes, e.g., tumor immunotherapy. In fact, pioneer studies of William Coley in cancer immunotherapy were based on administering bacterial products to cancer patients (64).

Vaccinia Virus

Live vaccinia virus was successfully used against smallpox until its eradication in 1977. Two observational studies carried out in Africa concluded that adults being smallpox vaccinated had significantly lower mortality risk, with a stronger effect observed in women than in men (65, 66). Since both studies were carried out when smallpox was already eradicated and, therefore, in

the absence of the targeted infection, the beneficial effects are necessarily non-specific. The capacity of a subset of NK cells to exhibit certain aspects of innate memory following infection with vaccinia virus was found by Gillard et al. in 2011. They demonstrated that this innate memory provides host protection against a subsequent systemic infection with a lethal dose of vaccinia virus, in some cases resulting in the complete clearance of detectable virus (67).

Influenza Vaccine

Trivalent live attenuated influenza vaccine has been shown to confer indirect protection from respiratory illness among children (68).

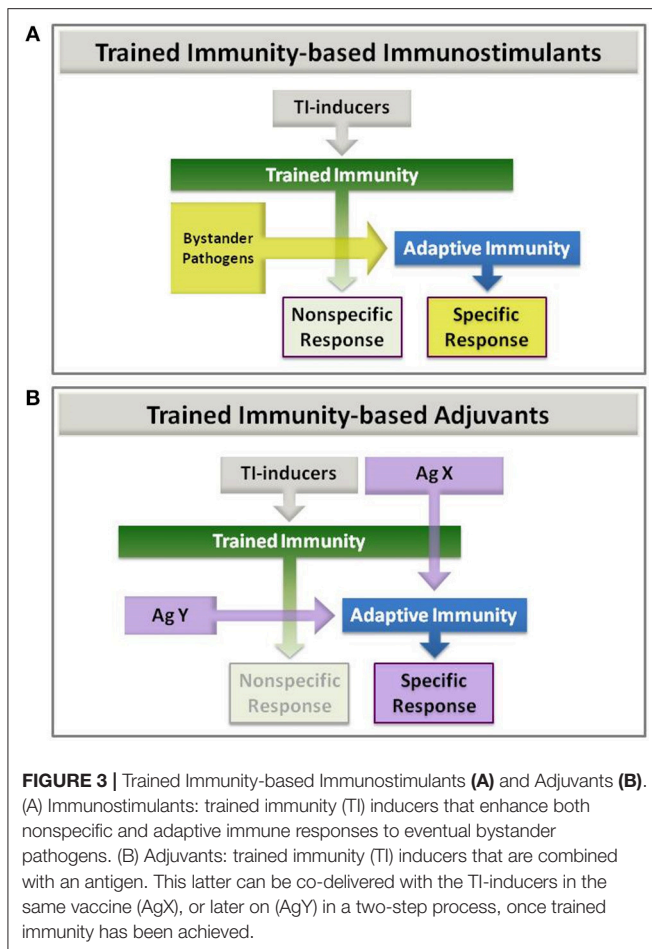
Respiratory syncytial virus (RSV) and influenza virus share common features, including innate immunity activation via PRRs, such as TLR3, TLR7, and retinoic acid-inducible gene I (RIG-I) (69, 70). The cold-adapted, live attenuated influenza vaccine (CAIV) has been shown to provide non-specific cross-protection against RSV in a murine model of infection (71). The results demonstrated that this vaccination induces local immune responses that provide a broad range of antiviral immunity, including protection against RSV, and that TLR3- and TLR7-mediated innate immunity plays an important role in protection against RSV (71).

Immunostimulants and Adjuvants as Putative TIbV

In addition to the examples described above, it is likely that other bacteria, fungi and viral preparations used as immunostimulants for different conditions might promote trained immunity if containing suitable inducers. In this regard, *Candida*-derived β -glucan is a paradigmatic example as it is a well-known inducer of trained immunity via dectin-1 (14). At this point, it should be noted that trained immunity-based immunostimulants might be considered within the TIbV concept because, under the umbrella of trained immunity, they may enhance innate and adaptive responses to bystander pathogens and their corresponding antigens (Figure 3A).

In 1986, Bistoni et al. demonstrated that systemic infection of mice with an avirulent *C. albicans* strain conferred protection not only against subsequent intravenous challenge with a pathogenic *C. albicans* strain but to *S. aureus* as well (72). More recently, it has been shown that protection from secondary lethal infection can be achieved with β -glucan and is dependent on epigenetic reprogramming linked to trained immunity (14). Although the immunostimulating effect of β -glucans is known for decades (73), the molecular mechanisms involved has started to be understood in the last few years (19, 21). The potential of oral β -glucan as “immune trainer” has been assayed in a pilot study in healthy volunteers (74). Innate immune responses were subsequently evaluated in peripheral blood mononuclear cells re-stimulated *in vitro* with *C. albicans*. However, the results showed a lack of cytokine production or microbicidal activity, which could be due to several reasons, including the dose and route of administration (74).

The immunostimulant OM-85 is a mixture of bacterial lysates for oral administration able to increase protection in



a murine model of respiratory viral infection, reducing viral load in the lung following experimental infection (75). It also reduced rhinovirus infection of lung epithelial cells (76) and had a protective role in models of viral/bacterial respiratory infections, reducing disease symptoms and improving survival (77). OM-85 has demonstrated clinical efficacy reducing the incidence, prevalence and/or duration of infections in children and adults (78–80). It is not known whether the mechanisms behind crossprotection of this and similar immunostimulants are dependent on trained immunity, but it is likely in analogy with the bacterial vaccines described above.

Another aspect of trained immunity-based immunostimulants is that they might be considered adjuvants when combined with other antigens to which an enhanced immune response is expected (Figure 3B; AgX). Thus, TIbV containing exogenous or chimeric antigens can be furnished. This possibility has been recently described for BCG being used as adjuvant for recombinant hepatitis B surface antigen vaccination (rHBsAg) (81). Even if the antigen may be administered in a second step once the trained immunity is induced (Figure 3B; AgY). In this composition the TIbV is split in two separated elements, the trained immunity-inducer and the antigen itself. In this sense, the influence of BCG on antibody and cytokine responses to human neonatal vaccination

has been described already (82). Either combination can be used for the development of novel vaccines with very specific but low antigenic molecules such as those synthetic peptides designed by reverse vaccinology.

CLINICAL APPLICATIONS OF TIBV

The development of TIbV may represent an advantage over conventional vaccines in certain settings. Under the trained immunity umbrella, broader and stronger immune responses may be expected, without the limitations of antigen specificity. Some applications include the following:

1. When conventional vaccines are not available for pathogens that produce recurrent infections, such as those responsible for many of the respiratory and urinary infections, the most common infections in humans (83, 84). At this point, it is interesting to remark that during the 1918 influenza pandemic outbreak (Spanish flu), bacterial vaccines were used for preventive purposes with quite surprising success given the viral nature of the infection. Although this observation may be explained by a mere prevention of secondary infections of *S. pneumoniae*, which was included in many of those vaccines (85), a non-mutually exclusive alternative might be that they were acting by inducing trained immunity, i.e., as TIbV, truly protecting from flu infection.
2. To prevent illness in which bacteria and virus co-infections have a role, such as asthma exacerbations (86).
3. When directed to pathogens with high mutation rates, such as influenza virus (87, 88). The broad spectrum of TIbV may circumvent the selective pressure of highly specific vaccines. This may also be applied to avoid the emergence of new bacteria strains upon conventional vaccination as occurs with vaccines directed to specific pneumococcal serotypes (89).
4. When used for preventive purposes for individuals susceptible to infections for which there are no vaccines, i.e., children and elderly population which are susceptible to mucosal infections (90, 91). Also in certain immunodeficiency states in which innate immunity is often preserved. In this sense, TIbV might be an achievable alternative to the use of broad range of antibiotics for preventive purposes in recurrent infections (92, 93).
5. Restoring immune responsiveness in clinical conditions associated with immune paralysis, such as severe sepsis and/or malignant processes (94).

In addition, the TIbV concept can be applied to the design of vaccines directed to the antigen of interest whether it is combined or not with the trained immunity stimuli. As mentioned above this concept is also broad and TIbV may be considered immunostimulants for endogenous (bystander) pathogens, or as a type of adjuvant for any antigen. While all these applications are of primary clinical interest, caution is needed about the potential deleterious function of trained immunity in patients suffering diseases characterized by excessive inflammation. This potential deleterious effect of trained immunity could apply to atherosclerosis (95), cardiovascular events (96), gout (97), and a

variety of autoimmune diseases and autoinflammatory disorders such as rheumatoid arthritis, systemic lupus erythematosus or hyper-IgD syndrome (98), where monocytes/macrophages show a detrimental trained immunity-like phenotype. Similarly, trained microglia has been linked to neurological disorders and stroke (99). Current knowledge, however, does not support such a deleterious role for TIBV: (a) Due to their nature, trained immunity is likely to have a transitory rather than a permanent effect, giving the system the required plasticity to avoid long-term potentially deleterious effects; (b) at least in two models (MV130 and MV140), TIBV have been shown to induce the production of the regulatory cytokine IL-10 by DCs and T cells, both *in vitro* and *in vivo* (47, 56); (c) MV130 has been shown to reduce recurrent infections in patients with rheumatoid arthritis without adverse effects (Candelas et al., manuscript in preparation), an observation also noted with other bacterial-derived immunostimulants (100, 101).

Many currently licensed vaccines consist of whole or inactivated pathogens; however, there has been a recent shift toward using simpler molecules such as highly purified antigens, recombinant, or synthetic peptides or DNA vaccines. Computational analysis of genetic sequences is now used for the prediction of just few T-cell epitopes fitting with most HLA molecules (102) as well as for antigen searching by the so-called reverse vaccinology (103). While these new generation of vaccines are focused to get a quite specific driven response, they are poorly immunogenic in absence of proper adjuvants. Thus, to confer protective immunity a strategy might be the combination of the adjuvant potential of trained immunity with the selected antigen epitopes. An important aspect to take into account is that it is not yet known whether or not all trained immunity stimuli produce the same functional behavior on innate immune cells with regard of driving T cells responses. As different PRR ligands may trigger different cell activation pathways with additive, synergies and opposite effects on key cell functions (44), it is likely that there are more than a single trained immunity functional program. Thus, the same antigen molecule can be eventually combined with different trained immunity stimuli for tailoring the better desired T cell response, like currently is being done with other adjuvants. In this regard, TLR8 agonists that mimic the immunomodulating effects of BCG, and enhance innate

and adaptive immune responses have been described recently (104).

FUTURE DIRECTIONS

Much knowledge is required in this field to successfully develop the potential of TIBV. Although their main advantage is that they act broadly on different pathogens and their potential as novel immunotherapy approach in both infectious, and even non-infectious immune related diseases such as cancer immunotherapy is obvious, their extent and limitations may depend on their composition. The pattern of response induced in the host innate immune cells by the specific training stimulus may also dictate the duration and interaction with ongoing specific responses. Despite the beneficial effects of TIBV inducing trained immunity as a host defense mechanism, any possible harmful effect in the induction and/or maintenance of autoimmune disorders cannot be ruled out, irrespective that current evidence does not support such a deleterious role. More studies will have to be carried out to fully understand the advantages and limitations of TIBV. Moreover, the indication of TIBV might consider several factors such as drug intake (105) or diet (106) that may affect the induction of trained immunity or its activation status, respectively. Finally, trained immunity-based stimulants and/or adjuvants widen the spectrum of the *in-silico* models for predicting immune responses. Thus, searching for new candidate combinations as trained immunity-based adjuvants for improved immunization purposes as TIBV is outlined as a novel and promising vaccine design.

AUTHOR CONTRIBUTIONS

SS-R, LC, and JS conducted literature searches, selected the studies, and wrote the manuscript. MN, DS, and ÓP critically contributed to the editing of the manuscript, final version and approval.

ACKNOWLEDGMENTS

We are grateful to Dr. Miguel Casanovas for helpful discussion. MN was supported by a Spinoza grant of the Netherlands Organization for Scientific Research.

REFERENCES

- Smith KA. Louis pasteur, the father of immunology? *Front Immunol.* (2012) 3:68. doi: 10.3389/fimmu.2012.00068
- Bourhy H, Perrot A, Cavaillon J-M. *Rabies. Vaccines: A Biography.* A. W. Arstentein. New York, NY: Springer-Verlag (2010). 73–85.
- Farber DL, Netea MG, Radbruch A, Rajewsky K, Zinkernagel RM. Immunological memory: lessons from the past and a look to the future. *Nat Rev Immunol.* (2016) 16:124–8. doi: 10.1038/nri.2016.13
- Pinna D, Corti D, Jarrossay D, Sallusto F, Lanzavecchia A. Clonal dissection of the human memory B-cell repertoire following infection and vaccination. *Eur J Immunol.* (2009) 39:1260–70. doi: 10.1002/eji.200839129
- Smith SA, Zhou Y, Olivarez NP, Broadwater AH, de Silva AM, Crowe JE Jr. Persistence of circulating memory B cell clones with potential for dengue virus disease enhancement for decades following infection. *J Virol.* (2012) 86:2665–75. doi: 10.1128/JVI.06335-11
- Cherry JD. Epidemic pertussis in 2012—the resurgence of a vaccine-preventable disease. *N Engl J Med.* (2012) 367:785–7. doi: 10.1056/NEJMp1209051
- Zinkernagel RM. What if protective immunity is antigen-driven and not due to so-called memory B and T cells? *Immunol Rev.* (2018) 283:238–46. doi: 10.1111/imr.12648
- Näslund C. *Resultats des Experiences de Vaccination par le BCG Poursuivies dans le Norrbotten (Suède) (Septembre 1927–Décembre 1931).* Vaccination Preventative de Tuberculose, Rapports et Documents. Paris: Institut Pasteur (1932).
- Aaby P, Roth A, Ravn H, Napirna BM, Rodrigues A, Lisse IM, et al. Randomized trial of BCG vaccination at birth to low-birth-weight children:

- beneficial nonspecific effects in the neonatal period? *J Infect Dis.* (2011) 204:245–52. doi: 10.1093/infdis/jir240
10. Benn CS, Netea MG, Selin LK, Aaby P. A small jab - a big effect: nonspecific immunomodulation by vaccines. *Trends Immunol.* (2013) 34:431–9. doi: 10.1016/j.it.2013.04.004
 11. Goodridge HS, Ahmed SS, Curtis N, Kollmann TR, Levy O, Netea MG, et al. Harnessing the beneficial heterologous effects of vaccination. *Nat Rev Immunol.* (2016) 16:392–400. doi: 10.1038/nri.2016.43
 12. Mayr A. Taking advantage of the positive side-effects of smallpox vaccination. *J Vet Med B Infect Dis Vet Public Health* (2004) 51:199–201. doi: 10.1111/j.1439-0450.2004.00763.x
 13. Netea MG, Quintin J, van der Meer JW. Trained immunity: a memory for innate host defense. *Cell Host Microbe* (2011) 9:355–61. doi: 10.1016/j.chom.2011.04.006
 14. Quintin J, Saeed S, Martens JHA, Giamarellos-Bourboulis EJ, Ifrim DC, Logie C, et al. *Candida albicans* infection affords protection against reinfection via functional reprogramming of monocytes. *Cell Host Microbe* (2012) 12:223–32. doi: 10.1016/j.chom.2012.06.006
 15. Sun JC, Ugolini S, Vivier E. Immunological memory within the innate immune system. *EMBO J.* (2014) 33:1295–303. doi: 10.1002/embj.201387651
 16. Gourbal B, Pinaud S, Beckers GJM, Van Der Meer JWM, Conrath U, Netea MG. Innate immune memory: an evolutionary perspective. *Immunol Rev.* (2018) 283:21–40. doi: 10.1111/immr.12647
 17. Quintin J, Cheng SC, van der Meer JW, Netea MG. Innate immune memory: towards a better understanding of host defense mechanisms. *Curr Opin Immunol.* (2014) 29:1–7. doi: 10.1016/j.coi.2014.02.006
 18. Saeed S, Quintin J, Kerstens HH, Rao NA, Aghajani-efah A, Matarese F, et al. Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. *Science* (2014) 345:1251086. doi: 10.1126/science.1251086
 19. Netea MG, Joosten LA, Latz E, Mills KH, Natoli G, Stunnenberg HG, et al. Trained immunity: a program of innate immune memory in health and disease. *Science* (2016) 352:aaf1098. doi: 10.1126/science.aaf1098
 20. Dominguez-Andres J, Joosten LA, Netea MG. Induction of innate immune memory: the role of cellular metabolism. *Curr Opin Immunol.* (2018) 56:10–6. doi: 10.1016/j.coi.2018.09.001
 21. Cheng SC, Quintin J, Cramer RA, Shephardson KM, Saeed S, Kumar V, et al. mTOR- and HIF-1 α -mediated aerobic glycolysis as metabolic basis for trained immunity. *Science* (2014) 345:1250684. doi: 10.1126/science.1250684
 22. Arts RJW, Carvalho A, La Rocca C, Palma C, Rodrigues F, Silvestre R, et al. Immunometabolic pathways in BCG-induced trained immunity. *Cell Rep.* (2016) 17:2562–71. doi: 10.1016/j.celrep.2016.11.011
 23. Gruenbacher G, Thurnher M. Mevalonate metabolism in cancer stemness and trained immunity. *Front Oncol.* (2018) 8:394. doi: 10.3389/fonc.2018.00394
 24. Kleinnijenhuis J, Quintin J, Preijers F, Joosten LA, Ifrim DC, Saeed S, et al. Bacille Calmette-Guérin induces NOD2-dependent nonspecific protection from reinfection via epigenetic reprogramming of monocytes. *Proc Natl Acad Sci USA.* (2012) 109:17537–42. doi: 10.1073/pnas.1202870109
 25. Arts RJW, Moorlag S, Novakovic B, Li Y, Wang SY, Oosting M, et al. BCG vaccination protects against experimental viral infection in humans through the induction of cytokines associated with trained immunity. *Cell Host Microbe* (2018b) 23:89–100e5. doi: 10.1016/j.chom.2017.12.010
 26. Kleinnijenhuis J, Quintin J, Preijers F, Benn CS, Joosten LA, Jacobs C, et al. Long-lasting effects of BCG vaccination on both heterologous Th1/Th17 responses and innate trained immunity. *J Innate Immun.* (2014) 6:152–8. doi: 10.1159/000355628
 27. Faure F, Jouve M, Lebhar-Peguillet I, Sadaka C, Sepulveda F, Lantz O, et al. Blood monocytes sample MelanA/MART1 antigen for long-lasting cross-presentation to CD8(+) T cells after differentiation into dendritic cells. *Int J Cancer* (2018) 142:133–44. doi: 10.1002/ijc.31037
 28. Kaufmann E, Sanz J, Dunn JL, Khan N, Mendonca LE, Pacis A, et al. BCG educates hematopoietic stem cells to generate protective innate immunity against tuberculosis. *Cell* (2018) 172:176–190e.19. doi: 10.1016/j.cell.2017.12.031
 29. Breyne K, Steenbrugge J, Demeyere K, Vanden Berghe T, Meyer E. Preconditioning with lipopolysaccharide or lipoteichoic acid protects against *Staphylococcus aureus* mammary infection in mice. *Front Immunol.* (2017) 8:833. doi: 10.3389/fimmu.2017.00833
 30. Krahenbuhl JL, Sharma SD, Ferraresi RW, Remington JS. Effects of muramyl dipeptide treatment on resistance to infection with *Toxoplasma gondii* in mice. *Infect Immun.* (1981) 31:716–22.
 31. Munoz N, Van Maele L, Marques JM, Rial A, Sirard JC, Chabalgoity JA. Mucosal administration of flagellin protects mice from *Streptococcus pneumoniae* lung infection. *Infect Immun.* (2010) 78:4226–33. doi: 10.1128/IAI.00224-10
 32. Zhang B, Chassaing B, Shi Z, Uchiyama R, Zhang Z, Denning TL, et al. Viral infection. Prevention and cure of rotavirus infection via TLR5/NLRC4-mediated production of IL-22 and IL-18. *Science* (2014) 346:861–65. doi: 10.1126/science.1256999
 33. Abdul-Careem MF, Firoz Mian M, Gillgrass AE, Chenoweth MJ, Barra NG, Chan T, et al. FimH, a TLR4 ligand, induces innate antiviral responses in the lung leading to protection against lethal influenza infection in mice. *Antiviral Res.* (2011) 92:346–55. doi: 10.1016/j.antiviral.2011.09.004
 34. Marakalala MJ, Williams DL, Hoving JC, Engstad R, Netea MG, Brown GD. Dectin-1 plays a redundant role in the immunomodulatory activities of beta-glucan-rich ligands *in vivo*. *Microbes Infect.* (2013) 15:511–5. doi: 10.1016/j.micinf.2013.03.002
 35. Rizzetto L, Ifrim DC, Moretti S, Tocci N, Cheng SC, Quintin J, et al. Fungal chitin induces trained immunity in human monocytes during cross-talk of the host with *Saccharomyces cerevisiae*. *J Biol Chem.* (2016) 291:7961–72. doi: 10.1074/jbc.M115.699645
 36. Ribes S, Meister T, Ott M, Redlich S, Janova H, Hanisch UK, et al. Intraperitoneal prophylaxis with CpG oligodeoxynucleotides protects neutropenic mice against intracerebral *Escherichia coli* K1 infection. *J Neuroinflammation* (2014) 11:14. doi: 10.1186/1742-2094-11-14
 37. Jiang T, Zhao H, Li XF, Deng YQ, Liu J, Xu LJ, et al. CpG oligodeoxynucleotides protect against the 2009 H1N1 pandemic influenza virus infection in a murine model. *Antiviral Res.* (2011) 89:124–6. doi: 10.1016/j.antiviral.2010.11.013
 38. Norton EB, Clements JD, Voss TG, Cardenas-Freytag L. Prophylactic administration of bacterially derived immunomodulators improves the outcome of influenza virus infection in a murine model. *J Virol.* (2010) 84:2983–95. doi: 10.1128/JVI.01805-09
 39. Schleicher U, Paduch K, Debus A, Obermeyer S, König T, Kling JC, et al. TNF-mediated restriction of arginase 1 expression in myeloid cells triggers type 2 no synthase activity at the site of infection. *Cell Rep.* (2016) 15:1062–75. doi: 10.1016/j.celrep.2016.04.001
 40. Mayer-Barber KD, Yan B. Clash of the cytokine titans: counter-regulation of interleukin-1 and type I interferon-mediated inflammatory responses. *Cell Mol Immunol.* (2017) 14:22–35. doi: 10.1038/cmi.2016.25
 41. Tejera-Alhambra M, Palomares O, Perez de Diego R, Diaz-Lezcano I, Sanchez-Ramon S. New biological insights in the immunomodulatory effects of mucosal polybacterial vaccines in clinical practice. *Curr Pharm Des.* (2016) 22:6283–93. doi: 10.2174/1381612822666160829143129
 42. Mourits VP, Wijkman JC, Joosten LA, Netea MG. Trained immunity as a novel therapeutic strategy. *Curr Opin Pharmacol.* (2018) 41:52–8. doi: 10.1016/j.coph.2018.04.007
 43. Tso GHW, Reales-Calderon JA, Pavelka N. The elusive anti-candida vaccine: lessons from the past and opportunities for the future. *Front Immunol.* (2018) 9:897. doi: 10.3389/fimmu.2018.00897
 44. Guy B. The perfect mix: recent progress in adjuvant research. *Nat Rev Microbiol.* (2007) 5:505–17. doi: 10.1038/nrmicro1681
 45. Blok BA, Arts RJ, van Crevel R, Benn CS, Netea MG. Trained innate immunity as underlying mechanism for the long-term, nonspecific effects of vaccines. *J Leukoc Biol.* (2015) 98:347–56. doi: 10.1189/jlb.5RI0315-096R
 46. de Bree C, Koeken V, Joosten LAB, Aaby P, Benn CS, van Crevel R, et al. Non-specific effects of vaccines: current evidence and potential implications. *Semin Immunol.* (2018). 39:35–43 doi: 10.1016/j.smim.2018.06.002
 47. Cirauqui C, Benito-Villalvilla C, Sanchez-Ramon S, Sirvent S, Diez-Rivero CM, Conejero L, et al. Human dendritic cells activated with MV130 induce Th1, Th17 and IL-10 responses via RIPK2 and MyD88 signalling pathways. *Eur J Immunol.* (2018) 48:180–93. doi: 10.1002/eji.201747024

48. Zanoni I, Tan Y, Di Gioia M, Broggi A, Ruan J, Shi J, et al. An endogenous caspase-11 ligand elicits interleukin-1 release from living dendritic cells. *Science* (2016) 352:1232–6. doi: 10.1126/science.aaf3036
49. Evavold CL, Kagan JC. How inflammasomes inform adaptive immunity. *J Mol Biol.* (2018) 430:217–37. doi: 10.1016/j.jmb.2017.09.019
50. Mitroulis I, Ruppova K, Wang B, Chen LS, Grzybek M, Grinenko T, et al. Modulation of myelopoiesis progenitors is an integral component of trained immunity. *Cell* (2018) 172:147–61.e12. doi: 10.1016/j.cell.2017.11.034
51. Alecsandru D, Valor L, Sanchez-Ramon S, Gil J, Carbone J, Navarro J, et al. Sublingual therapeutic immunization with a polyvalent bacterial preparation in patients with recurrent respiratory infections: immunomodulatory effect on antigen-specific memory CD4+ T cells and impact on clinical outcome. *Clin Exp Immunol.* (2011) 164:100–7. doi: 10.1111/j.1365-2249.2011.04320.x
52. Lorenzo-Gomez MF, Padilla-Fernandez B, Garcia-Criado FJ, Miron-Canelo JA, Gil-Vicente A, Nieto-Huertos A, et al. Evaluation of a therapeutic vaccine for the prevention of recurrent urinary tract infections versus prophylactic treatment with antibiotics. *Int Urogynecol J.* (2013) 24:127–34. doi: 10.1007/s00192-012-1853-5
53. Lorenzo-Gomez MF, Padilla-Fernandez B, Garcia-Cenador MB, Virseda-Rodriguez AJ, Martin-Garcia I, Sanchez-Escudero A, et al. Comparison of sublingual therapeutic vaccine with antibiotics for the prophylaxis of recurrent urinary tract infections. *Front Cell Infect Microbiol.* (2015) 5:50. doi: 10.3389/fcimb.2015.00050
54. Sanchez-Ramon S, Perez de Diego R, Dieli-Crimi R, Subiza JL. Extending the clinical horizons of mucosal bacterial vaccines: current evidence and future prospects. *Curr Drug Targets* (2014) 15:1132–43. doi: 10.2174/1389450115666141020160705
55. Yang B, Foley S. First experience in the UK of treating women with recurrent urinary tract infections with the bacterial vaccine Uromune(R). *BJU Int.* (2018) 121:289–92. doi: 10.1111/bju.14067
56. Benito-Villalvilla C, Cirauqui C, Diez-Rivero CM, Casanovas M, Subiza JL, Palomares O. MV140, a sublingual polyvalent bacterial preparation to treat recurrent urinary tract infections, licenses human dendritic cells for generating Th1, Th17, and IL-10 responses via Syk and MyD88. *Mucosal Immunol.* (2017) 10:924–35. doi: 10.1038/mi.2016.112
57. Birk NM, Nissen TN, Kjaergaard J, Hartling HJ, Thostesen LM, Kofoed PE, et al. Effects of Bacillus Calmette-Guerin (BCG) vaccination at birth on T and B lymphocyte subsets: results from a clinical randomized trial. *Sci Rep.* (2017) 7:12398. doi: 10.1038/s41598-017-11601-6
58. Jensen KJ, Benn CS, van Crevel R. Unravelling the nature of non-specific effects of vaccines-A challenge for innate immunologists. *Semin Immunol.* (2016) 28:377–83. doi: 10.1016/j.smim.2016.05.005
59. Kjaergaard J, Birk NM, Nissen TN, Thostesen LM, Pihl GT, Benn CS, et al. Nonspecific effect of BCG vaccination at birth on early childhood infections: a randomized, clinical multicenter trial. *Pediatr Res.* (2016) 80:681–5. doi: 10.1038/pr.2016.142
60. de Bree C, Marijnissen RJ, Kel JM, Rosendahl Huber SK, Aaby P, Benn CS, et al. (2018). Bacillus Calmette-guerin-induced trained immunity is not protective for experimental influenza A/Anhui/1/2013 (H7N9) infection in mice. *Front Immunol.* 9:869. doi: 10.3389/fimmu.2018.00869
61. Kawai K, Miyazaki J, Joraku A, Nishiyama H, Akaza H. Bacillus Calmette-Guerin (BCG) immunotherapy for bladder cancer: current understanding and perspectives on engineered BCG vaccine. *Cancer Sci.* (2013) 104:22–7. doi: 10.1111/cas.12075
62. Buffen K, Oosting M, Quintin J, Ng A, Kleinnijenhuis J, Kumar V, et al. Autophagy controls BCG-induced trained immunity and the response to intravesical BCG therapy for bladder cancer. *PLoS Pathog.* (2014) 10:e1004485. doi: 10.1371/journal.ppat.1004485
63. Pieraerts C, Martin V, Jichlinski P, Nardelli-Haeffliger D, Derre L. Detection of functional antigen-specific T cells from urine of non-muscle invasive bladder cancer patients. *Oncoimmunology* (2012) 1:694–8. doi: 10.4161/onci.20526
64. Chamberlain RS, Kaufman H. Innovations and strategies for the development of anticancer vaccines. *Expert Opin Pharmacother.* (2000) 1:603–14. doi: 10.1517/14656566.1.4.603
65. Aaby P, Gustafson P, Roth A, Rodrigues A, Fernandes M, Sodemann M, et al. Vaccinia scars associated with better survival for adults. An observational study from Guinea-Bissau. *Vaccine* (2006) 24:5718–25. doi: 10.1016/j.vaccine.2006.04.045
66. Jensen ML, Dave S, Schim van der Loeff M, da Costa C, Vincent T, Leliggowicz A, et al. Vaccinia scars associated with improved survival among adults in rural Guinea-Bissau. *PLoS ONE* (2006) 1:e101. doi: 10.1371/journal.pone.0000101
67. Gillard GO, Bivas-Benita M, Hovav AH, Grandpre LE, Panas MW, Seaman MS, et al. Thy1+ NK [corrected] cells from vaccinia virus-primed mice confer protection against vaccinia virus challenge in the absence of adaptive lymphocytes. *PLoS Pathog.* (2011) 7:e1002141. doi: 10.1371/journal.ppat.1002141
68. Piedra PA, Gaglani MJ, Kozinetz CA, Herschler GB, Fewless C, Harvey D, et al. Trivalent live attenuated intranasal influenza vaccine administered during the 2003-2004 influenza type A (H3N2) outbreak provided immediate, direct, and indirect protection in children. *Pediatrics* (2007) 120:e553–64. doi: 10.1542/peds.2006-2836
69. Diebold SS, Kaisho T, Hemmi H, Akira S, Reis e Sousa C. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science* (2004) 303:1529–31. doi: 10.1126/science.1093616
70. Guillot L, Le Goffic R, Bloch S, Escricou N, Akira S, Chignard M, et al. Involvement of toll-like receptor 3 in the immune response of lung epithelial cells to double-stranded RNA and influenza A virus. *J Biol Chem.* (2005) 280:5571–80. doi: 10.1074/jbc.M410592200
71. Lee YJ, Lee JY, Jang YH, Seo SU, Chang J, Seong BL. Non-specific effect of vaccines: immediate protection against respiratory syncytial virus infection by a live attenuated influenza vaccine. *Front Microbiol.* (2018) 9:83. doi: 10.3389/fmicb.2018.00083
72. Bistoni F, Vecchiarelli A, Cenci E, Puccetti P, Marconi P, Cassone A. Evidence for macrophage-mediated protection against lethal *Candida albicans* infection. *Infect Immun.* (1986) 51:668–74.
73. Brown GD, Gordon S. Fungal beta-glucans and mammalian immunity. *Immunity* (2003) 19:311–5. doi: 10.1016/S1074-7613(03)00233-4
74. Leentjens J, Quintin J, Gerretsen J, Kox M, Pickkers P, Netea MG. The effects of orally administered Beta-glucan on innate immune responses in humans, a randomized open-label intervention pilot-study. *PLoS ONE* (2014) 9:e108794. doi: 10.1371/journal.pone.0108794
75. Pasquali C, Salami O, Taneja M, Gollwitzer ES, Trompette A, Pattaroni C, et al. Enhanced mucosal antibody production and protection against respiratory infections following an orally administered bacterial extract. *Front Med.* (2014) 1:41. doi: 10.3389/fmed.2014.00041
76. Roth M, Pasquali C, Stolz D, Tamm M. Broncho Vaxom (OM-85) modulates rhinovirus docking proteins on human airway epithelial cells via Erk1/2 mitogen activated protein kinase and cAMP. *PLoS ONE* (2017) 12:e0188010. doi: 10.1371/journal.pone.0188010
77. Bessler WG, Vor dem Esche U, Masihi N. The bacterial extract OM-85 BV protects mice against influenza and Salmonella infection. *Int Immunopharmacol.* (2010) 10:1086–90. doi: 10.1016/j.intimp.2010.06.009
78. Collet JP, Shapiro P, Ernst P, Renzi T, Ducruet T, Robinson A. Effects of an immunostimulating agent on acute exacerbations and hospitalizations in patients with chronic obstructive pulmonary disease. The PARI-IS study steering committee and research group prevention of acute respiratory infection by an immunostimulant. *Am J Respir Crit Care Med.* (1997) 156:1719–24. doi: 10.1164/ajrcrm.156.6.9612096
79. Gutierrez-Tarango MD, Berber A. Safety and efficacy of two courses of OM-85 BV in the prevention of respiratory tract infections in children during 12 months. *Chest* (2001) 119:1742–8. doi: 10.1378/chest.119.6.1742
80. Razi CH, Harmanici K, Abaci A, Ozdemir O, Hizli S, Renda R, et al. The immunostimulant OM-85 BV prevents wheezing attacks in preschool children. *J Allergy Clin Immunol.* (2010) 126:763–9. doi: 10.1016/j.jaci.2010.07.038
81. Scheid A, Borriello F, Pietrasanta C, Christou H, Diray-Arce J, Pettengill MA, et al. Adjuvant effect of bacille calmette-guerin on hepatitis B vaccine immunogenicity in the preterm and term newborn. *Front Immunol.* (2018) 9:29. doi: 10.3389/fimmu.2018.00029
82. Ota MO, Vekemans J, Schlegel-Haueter SE, Fielding K, Sanneh M, Kidd M, et al. Influence of *Mycobacterium bovis* bacillus Calmette-Guerin on antibody and cytokine responses to human neonatal vaccination. *J Immunol.* (2002) 168:919–25. doi: 10.4049/jimmunol.168.2.919

83. Mizgerd JP. Lung infection—a public health priority. *PLoS Med.* (2006) 3:e76. doi: 10.1371/journal.pmed.0030076
84. Foxman B. The epidemiology of urinary tract infection. *Nat Rev Urol.* (2010) 7:653–60. doi: 10.1038/nrurol.2010.190
85. Chien YW, Klugman KP, Morens DM. Efficacy of whole-cell killed bacterial vaccines in preventing pneumonia and death during the 1918 influenza pandemic. *J Infect Dis.* (2010) 202:1639–48. doi: 10.1086/657144
86. Kloepper KM, Lee WM, Pappas TE, Kang TJ, Vrtis RF, Evans MD, et al. Detection of pathogenic bacteria during rhinovirus infection is associated with increased respiratory symptoms and asthma exacerbations. *J Allergy Clin Immunol.* (2014) 133:1301–7, 1307.e1–3. doi: 10.1016/j.jaci.2014.02.030
87. Ndung'u T, Weiss RA. On HIV diversity. *AIDS* (2012) 26:1255–60. doi: 10.1097/QAD.0b013e32835461b5
88. Kumar A, Meldgaard TS, Bertholet S. Novel platforms for the development of a universal influenza vaccine. *Front Immunol.* (2018) 9:600. doi: 10.3389/fimmu.2018.00600
89. Pichichero ME. Pneumococcal whole-cell and protein-based vaccines: changing the paradigm. *Expert Rev Vaccines* (2017) 16:1181–90. doi: 10.1080/14760584.2017.1393335
90. Kline KA, Bowdish DM. Infection in an aging population. *Curr Opin Microbiol.* (2016) 29:63–7. doi: 10.1016/j.mib.2015.11.003
91. Esposito S, Soto-Martinez ME, Feleszko W, Jones MH, Shen KL, Schaad UB. Nonspecific immunomodulators for recurrent respiratory tract infections, wheezing and asthma in children: a systematic review of mechanistic and clinical evidence. *Curr Opin Allergy Clin Immunol.* (2018) 18:198–209. doi: 10.1097/ACI.0000000000000433
92. Wilson R, Sethi S, Anzueto A, Miravittles M. Antibiotics for treatment and prevention of exacerbations of chronic obstructive pulmonary disease. *J Infect.* (2013) 67:497–515. doi: 10.1016/j.jinf.2013.08.010
93. Abzug MJ. Acute sinusitis in children: do antibiotics have any role? *J Infect.* (2014) 68(Suppl 1):S33–7. doi: 10.1016/j.jinf.2013.09.012
94. Novakovic B, Habibi E, Wang SY, Arts RJW, Davar R, Megchelenbrink W, et al. Beta-glucan reverses the epigenetic state of LPS-induced immunological tolerance. *Cell* (2016) 167:1354–68.e14. doi: 10.1016/j.cell.2016.09.034
95. Leentjens J, Bekkering S, Joosten LAB, Netea MG, Burgner DP, Riksen NP. Trained innate immunity as a novel mechanism linking infection and the development of atherosclerosis. *Circ Res.* (2018) 122:664–9. doi: 10.1161/CIRCRESAHA.117.312465
96. Hoogveen RM, Nahrendorf M, Riksen NP, Netea MG, de Winther MPJ, Lutgens E, et al. Monocyte and haematopoietic progenitor reprogramming as common mechanism underlying chronic inflammatory and cardiovascular diseases. *Eur Heart J.* (2017) 39:3521–7. doi: 10.1093/eurheartj/ehx581
97. Crisan TO, Netea MG, Joosten LA. Innate immune memory: Implications for host responses to damage-associated molecular patterns. *Eur J Immunol.* (2016) 46:817–28. doi: 10.1002/eji.201545497
98. Arts RJW, Joosten LAB, Netea MG. The potential role of trained immunity in autoimmune and autoinflammatory disorders. *Front Immunol.* (2018a) 9:298. doi: 10.3389/fimmu.2018.00298
99. Wendeln AC, Degenhardt K, Kaurani L, Gertig M, Ulas T, Jain G, et al. Innate immune memory in the brain shapes neurological disease hallmarks. *Nature* (2018) 556:332–8. doi: 10.1038/s41586-018-0023-4
100. Chiavaroli C, Moore A. An hypothesis to link the opposing immunological effects induced by the bacterial lysate OM-89 in urinary tract infection and rheumatoid arthritis. *BioDrugs* (2006) 20:141–9. doi: 10.2165/00063030-200620030-00001
101. Toussiro E, Robinet E, Saas P, Chabod J, Auge B, Cozma G, et al. Bacterial extract (OM-89) specific and non specific immunomodulation in rheumatoid arthritis patients. *Autoimmunity* (2006) 39:299–306. doi: 10.1080/08916930600738425
102. Damfo SA, Reche P, Gatherer D, Flower DR. *In silico* design of knowledge-based *Plasmodium falciparum* epitope ensemble vaccines. *J Mol Graph Model.* (2017) 78:195–205. doi: 10.1016/j.jmgm.2017.10.004
103. Hegde NR, Gauthami S, Sampath Kumar HM, Bayry J. The use of databases, data mining and immunoinformatics in vaccinology: where are we? *Expert Opin Drug Discov.* (2018) 13:117–30. doi: 10.1080/17460441.2018.1413088
104. Dowling DJ, Scott EA, Scheid A, Bergelson I, Joshi S, Pietrasanta C, et al. Toll-like receptor 8 agonist nanoparticles mimic immunomodulating effects of the live BCG vaccine and enhance neonatal innate and adaptive immune responses. *J Allergy Clin Immunol.* (2017) 140:1339–50. doi: 10.1016/j.jaci.2016.12.985
105. Bekkering S, Arts RJW, Novakovic B, Kourtzelis IC, van der Heijden CDCC, Li Y, et al. Metabolic induction of trained immunity through the mevalonate pathway. *Cell* (2018) 172:135–46.e9. doi: 10.1016/j.cell.2017.11.025
106. Christ A, Gunther P, Lauterbach MAR, Duewell P, Biswas D, Pelka K, et al. Western diet triggers NLRP3-dependent innate immune reprogramming. *Cell* (2018) 172:162–175.e14. doi: 10.1016/j.cell.2017.12.013

Conflict of Interest Statement: JS is the CEO of Inmunotek SL, a pharmaceutical company that manufactures bacterial vaccines. LC is an employee of Inmunotek. SS-R, DS, and OP have received research grants from Inmunotek.

The handling Editor declared a shared affiliation, though no other collaboration, with two of the authors OP and SS-R.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Sánchez-Ramón, Conejero, Netea, Sancho, Palomares and Subiza. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genome-Based Approach Delivers Vaccine Candidates Against *Pseudomonas aeruginosa*

Irene Bianconi^{1†}, Beatriz Alcalá-Franco^{1†}, Maria Scarselli^{2†}, Mattia Dalsass^{2,3}, Scilla Buccato², Annalisa Colaprico², Sara Marchi², Vega Masignani^{2‡} and Alessandra Bragonzi^{1*‡}

¹ Infection and Cystic Fibrosis Unit, Division of Immunology, Transplantation and Infectious Diseases, IRCCS San Raffaele Scientific Institute, Milan, Italy, ² GSK, Siena, Italy, ³ Dipartimento di Scienze Cliniche e Biologiche, Università degli Studi di Torino, Turin, Italy

OPEN ACCESS

Edited by:

Pedro A. Reche,
Complutense University of Madrid,
Spain

Reviewed by:

Paola Massari,
Tufts University School of Medicine,
United States
Giampiero Pietroccola,
University of Pavia, Italy

*Correspondence:

Alessandra Bragonzi
bragonzi.alessandra@hsr.it

[†]These authors have contributed
equally to this work and share first
authorship

[‡]These authors have contributed
equally to this work and share senior
authorship

Specialty section:

This article was submitted to
Vaccines and Molecular Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 17 September 2018

Accepted: 06 December 2018

Published: 09 January 2019

Citation:

Bianconi I, Alcalá-Franco B,
Scarselli M, Dalsass M, Buccato S,
Colaprico A, Marchi S, Masignani V
and Bragonzi A (2019)
Genome-Based Approach Delivers
Vaccine Candidates Against
Pseudomonas aeruginosa.
Front. Immunol. 9:3021.
doi: 10.3389/fimmu.2018.03021

High incidence, severity and increasing antibiotic resistance characterize *Pseudomonas aeruginosa* infections, highlighting the need for new therapeutic options. Vaccination strategies to prevent or limit *P. aeruginosa* infections represent a rational approach to positively impact the clinical outcome of risk patients; nevertheless this bacterium remains a challenging vaccine target. To identify novel vaccine candidates, we started from the genome sequence analysis of the *P. aeruginosa* reference strain PAO1 exploring the reverse vaccinology approach integrated with additional bioinformatic tools. The bioinformatic approaches resulted in the selection of 52 potential antigens. These vaccine candidates were conserved in *P. aeruginosa* genomes from different origin and among strains isolated longitudinally from cystic fibrosis patients. To assess the immune-protection of single or antigens combination against *P. aeruginosa* infection, a vaccination protocol was established in murine model of acute respiratory infection. Combinations of selected candidates, rather than single antigens, effectively controlled *P. aeruginosa* infection in the *in vivo* model of murine pneumonia. Five combinations were capable of significantly increase survival rate among challenged mice and all included PA5340, a hypothetical protein exclusively present in *P. aeruginosa*. PA5340 combined with PA3526-MotY gave the maximum protection. Both proteins were surface exposed by immunofluorescence and triggered a specific immune response. Combination of these two protein antigens could represent a potential vaccine to prevent *P. aeruginosa* infection.

Keywords: *Pseudomonas aeruginosa*, reverse vaccinology, vaccine, respiratory infection, mouse model

INTRODUCTION

P. aeruginosa infections are among the most severe public health issues. This opportunistic bacterium belongs to the multi-drug resistant (MDR) ESCAPE pathogens, along with *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, and *Enterobacter*. According to data from Centers for Disease Control, *P. aeruginosa* is responsible for millions of infections each year in the community, 10–15% of all healthcare-associated infections, with more than 300,000 cases annually in the EU, USA and Japan (1). Patients hospitalized in intensive care units (ICU) run a high risk of acquiring *P. aeruginosa* as they

may develop ventilator-associated pneumonia (VAP) and sepsis (2–4). Other risk groups are patients with a compromised immune system, either from immunosuppressive therapies and underlying diseases such as cancer, AIDS or hereditary cystic fibrosis (CF). This high prevalence is partly due to the vast arsenal of virulence factors that facilitates acute infections and the propensity of *P. aeruginosa* to form highly structured biofilm communities that cause chronic infections (5).

Taking the severity of the illness into account, current treatment guidelines for the management of bacterial infection recommend single antibiotic or combination therapy. Despite the wide arsenal of drugs for *P. aeruginosa* infections available on the market, inefficacy of these treatments is commonplace. Resistance rapidly emerges, usually linked to intrinsic bacterial resistance mechanisms, development of new antibiotic resistance, and/or limited penetration of antibiotics into biofilms (3, 6). Development of antibiotics with a novel mode of action and/or alternative therapies remains an urgent need for patients. Antibacterial agents launched in recent decades were modifications of existing molecules; the development of entirely new classes of antibiotics has been largely abandoned (7). Immunotherapy for preventing pulmonary infection has also been tested (8), but clinical efficacy has been disappointing (9, 10). Immunization strategies do not cover *P. aeruginosa* infection in healthcare practices.

In recent years, remarkable progress has been made in the identification of *P. aeruginosa* virulence factors and their variations among different infection processes. It has been more accurately recognized that *P. aeruginosa* is an antigenically variable microorganism that adapts easily to different growth conditions and escapes host immune recognition. The high variability of the proteins among different *P. aeruginosa* strains and within the same strain, grown in diverse environmental conditions, may represent a serious obstacle to the development of a globally effective anti-*P. aeruginosa* vaccine (10). So far, *P. aeruginosa* vaccine candidates have been found by classical approach—by identifying more abundant surface proteins and oligosaccharides or by selecting specific virulence factors, according to their relevance in the disease outcome. Integrated genomics and proteomics approaches have been recently used to predict vaccine candidates against *P. aeruginosa* (11). Although several vaccine formulations have been tested clinically, none has been licensed (10, 12). *P. aeruginosa* vaccines tested so far in humans consisted of antigens targeting single rather than multiple virulence mechanisms—OprF–OprI fusion (13), flagella (14), O antigen-conjugated vaccines (15), high molecular weight alginate (16). Further success in *P. aeruginosa* vaccine development may require a different approach, including bacterial genome evaluation to identify novel antigen combinations potentially addressing multiple virulence mechanisms, such as initial bacterial colonization, immune evasion, colony aggregation and cytotoxicity.

During the past two decades, reverse vaccinology has revolutionized the approach to vaccine research (17, 18), ultimately leading to the development of new generation vaccines based on antigens previously unrecognized by other approaches. One is 4CMenB, the first universal vaccine against

serogroup B *Neisseria meningitidis* (MenB), now licensed in several countries worldwide (19, 20). Reverse vaccinology aims at identifying surface exposed proteins, ideally playing a relevant role in pathogenesis, which can serve as targets of the host immune system. This approach has not yet been implemented for *P. aeruginosa*. In this study, reverse vaccinology approach was combined with advanced genomic technologies to select new protein antigens against *P. aeruginosa*. We report that combinations of selected candidates, more than single antigens, effectively control *P. aeruginosa* infections in a mouse model of acute pneumonia.

RESULTS

P. aeruginosa Antigens Selection by Genome-Wide Screening

Among 5,570 open reading frames (ORFs) encoded by *P. aeruginosa* PAO1 strain (5), we predicted a total of 2,430 surface or membrane-associated proteins by using high throughput bioinformatics localization prediction tools. In particular, subcellular localization was predicted by PsortB software. In the case of predicted non-cytoplasmic polypeptides the presence of signal peptide and localization of cleavage site were predicted with SignalP. N-terminal signatures predictive of lipoproteins were identified by using the LipoP server and putative transmembrane regions were predicted with Tmpred. Among them, 307 were classified as outer membrane proteins or lipoproteins, 583 as periplasmic proteins, and 2,109 as inner membrane proteins. The remaining 2,562 ORFs were predicted to encode cytoplasmic proteins. Since inner membrane proteins are barely exposed on the outside of the bacterium and are difficult to express and purify, all were discarded from selection except those with sequence similarities to known virulence factors or extracellular proteins from other bacterial pathogens. The final selection totaled 950 ORFs (Figure 1).

In a second step of prioritization and to avoid selection of potential self-antigens, we excluded all proteins containing domains with sequence similarity (E -value $> 1e-10$) to human and/or mouse proteins, narrowing the selection to 824 proteins. To identify proteins more directly related to *P. aeruginosa* pathogenesis and fitness, and to avoid widespread bacterial housekeeping factors, sequence comparison to *E. coli* K12 whole proteome was performed; we discarded *P. aeruginosa* proteins sharing more than 40% sequence similarity over at least 70% of the length of the *E. coli* counterpart. As sequence conservation is highly desirable for broad-spectrum vaccine candidates, a comparative analysis was performed with the genome sequence of seven published *P. aeruginosa* strains including clinical isolates of different origins (PA14, LESB58, PA7, 2192, C3719, PACS2, RP73); only proteins belonging to the core genome were kept. Candidate selection was also refined by removing short peptides (< 150 aa long), eventually leading to a total of 531 hits. To further prioritize the candidates and reduce the final pool of proteins for experimental testing, a PSI-Blast analysis was conducted; results were manually curated to remove any residual protein putatively involved in intermediate metabolism, DNA synthesis, translation

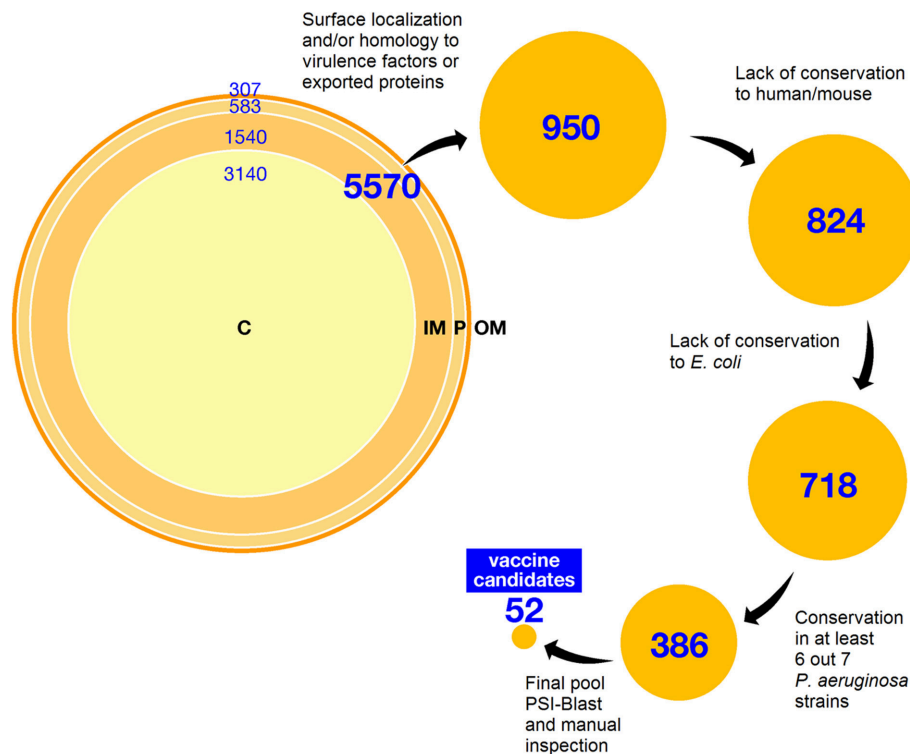


FIGURE 1 | Schematic workflow of antigen selection *in silico*. The complete PAO1 proteome has been analyzed as described in the text. Distribution of ORFs according to predicted sub-cellular localization using Psort is summarized at the top. Below, selection steps succession is reported, indicating the number of candidates (hits) selected after each step. C: cytoplasm; IN: inner membrane; P: periplasm; OM: outer membrane.

and repair, protein synthesis and transport, and more generally, in any cellular process confined to the bacterial cytoplasm. The final pool of *in silico* selected candidates included 52 antigens—31 proteins of known and 21 of unknown functions (**Figure 1** and **Table S1**). The presence of well-known virulence factors like ExoA and ExoT, as well as relevant outer membrane proteins like OprF and OprH in the final list of candidates confirmed the reliability of the selection strategy.

When tested against an extended panel of 104 *P. aeruginosa* complete genomes, it emerged that all candidates share a mean identity/coverage ratio ranging from 0.78 to 1.00 (**Figure S1**) confirming that a vast proportion of epitopes potentially presented by each candidate to the immune system is conserved across the natural *P. aeruginosa* population.

Evaluation of Candidates in a Murine Model of Pneumonia

Of 52 *P. aeruginosa* vaccine candidates (**Table S1**), 30 ORFs (57.7%) were successfully expressed in *E. coli* BL21 as His-tag fusions. OprF-OprI fusion, designed according to the known construct used in the recent clinical trial (21), was included in this study. Ability of these antigens to protect against *P. aeruginosa* infection was tested in a mouse model of acute pneumonia (22). C57Bl/6 mice were immunized intraperitoneally (i.p.) with 10 µg of each protein formulated with aluminum hydroxide (Alum) as adjuvant at 0, 21, and 35 days. At day 50, mice were

challenged intratracheally (i.t.) with 5×10^6 CFU/lung of *P. aeruginosa* reference strain PAO1 and monitored twice a day for 1 week for health parameters indicative of animal wellness. In this model, all mice immunized with Alum alone (negative control group) consistently showed symptoms of a severe clinical disease and died within 48 h, whereas mice immunized with whole cell inactivated PAO1 bacteria were consistently protected by homologous challenge in a dose-dependent manner. Of the 30 antigens tested, 10 showed a modest increase in survival compared to the negative control (up to 20% at day 5) and were investigated further (**Table 1** and **Figure 2**). The other 20 proteins tested did not differ substantially from the negative control and were discarded (data not shown). Survival of mice vaccinated with OprF-OprI fusion (10.7%) was comparable to that observed by vaccination with the 10 selected proteins.

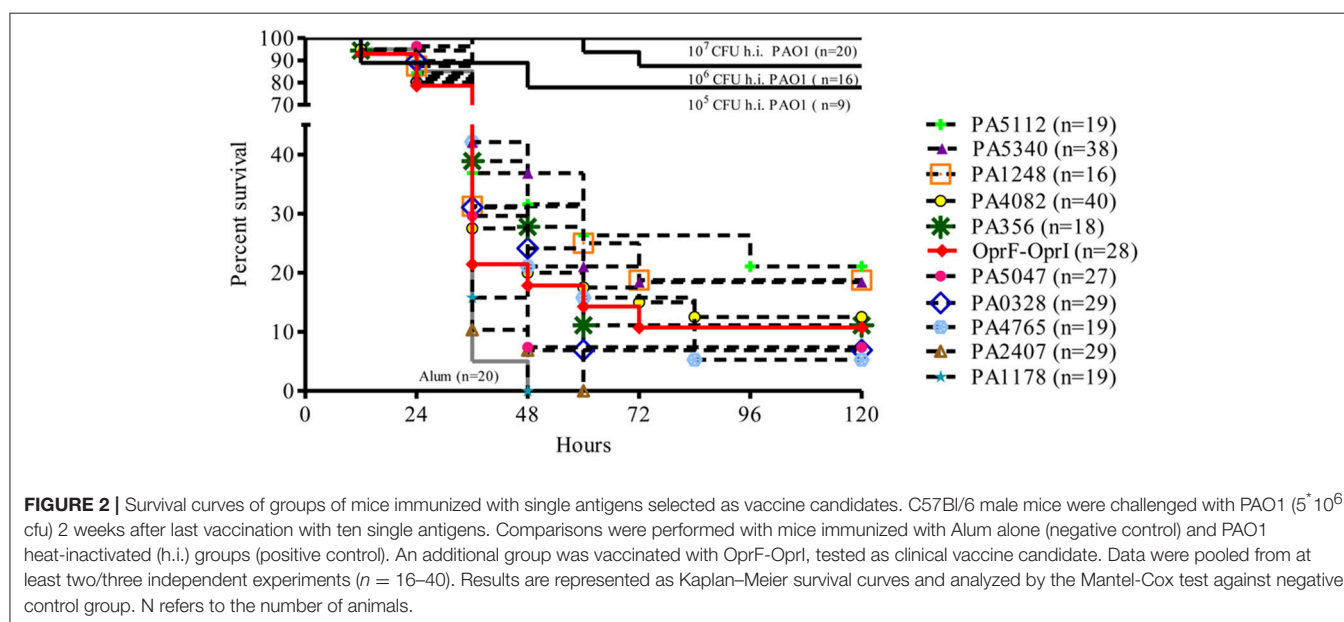
Evaluation of Combinations of Candidates in a Murine Model of Pneumonia

In order to increase the survival rate of vaccinated mice, proteins conferring higher protection were pooled in group of two, and 22 combinations were tested. Five combinations, all containing the antigen candidate PA5340, were the most promising (PA5340+PA1178-OprH, PA5340+PA3526-motY, PA5340+PA5112-EstA, PA5340+PA5047, PA5340+PA0328-AaaA). These combinations showed a significant increase in both survival curves (Mantel-Cox test $p < 0.0002$, 0.0019, 0.0027,

TABLE 1 | Top line vaccine candidates of *P. aeruginosa*.

Locus_tag	Protein annotation	Surface exposure ^(a)	Antigenic potential ^(b)	Sequence conservation ^(c)	Sequence conservation ^(d)
PA0328	arginine-specific autotransporter AaaA	Yes	Yes	99.3% ± 0.2%	99.2% ± 0.3%
PA1178	outer membrane protein OprH precursor	Yes	Yes	100% ± 0%	99.9% ± 0.2%
PA1248	outer membrane protein AprF precursor	Yes	Yes	99.6% ± 0.3%	99.2% ± 4.6%
PA2407	putative adhesion protein FpvC	Yes	Yes	100% ± 0%	98.3% ± 10.4%
PA3526	outer membrane protein precursor MotY	Yes	Yes	99.1% ± 0.8%	99.7% ± 0.5%
PA4082	adhesive protein CupB5	Yes	Yes	98.8% ± 0.5%	98.5% ± 2.9%
PA4765	outer membrane lipoprotein OmlA precursor	Yes	Yes	99.7% ± 0.3%	98.9% ± 7.2%
PA5047	putative Zn-dependent protease	Yes	Yes	99.9% ± 0.1%	99.2% ± 7.1%
PA5112	esterase EstA	No	Yes	99.8% ± 0.2%	99.1% ± 7.2%
PA5340	hypothetical protein	Yes	Yes	98.8% ± 0.3%	99.2% ± 0.3%

^aSurface exposure suggested by immunofluorescence microscopy co-localization (see text). ^bEvaluated by Western Blot against recombinant proteins, *P. aeruginosa* strain PAO1 and clinical isolate MDR-RP73 (see text). ^cSequence conservation expressed as mean percentage of amino acid identity ± standard deviation calculated among a collection of CF clinical strains ^cand on the public collection of 104 completed *P. aeruginosa* genomes available in GenBank (see text)^d.



0.015, and 0.015, respectively) and mean survival time (one-way ANOVA p -value < 0.01) when compared with negative controls (Figure 3 and Figure S2). The best antigens combination was PA5340+PA3526-MotY, with survival increased up to 50%. Three combinations described above, PA5340+PA3526-motY, PA5340+PA5112-EstA, and PA5340+PA0328-AaaA, increased survival significantly when compared with OprF-OprI (Mantel-Cox test 0.0091, 0.0009, and 0.012, respectively). Interestingly, an increase in survival rate was also observed (though not statistically significant) when OprF-OprI was combined with PA5340, going from 10.7% of the fusion alone to 40% when tested in combination (Mantel-Cox test 0.058).

In vitro Characterization of Selected Antigens

To characterize antigenic potential and cellular localization of selected antigens (PA1178-OprH, PA1248-AprF, PA5112-EstA,

PA0328-AaaA, PA2407-FpvC, PA3526-MotY, PA4082-CupB5, PA4765-OmlA, PA5047, and PA5340) the antisera obtained immunizing with recombinant proteins were used in Western Blot and immunofluorescence microscopy (Table 1 and Figure 4). All the antisera recognized the recombinant proteins, the homologous *P. aeruginosa* strain PAO1 and the heterologous clinical isolate MDR-RP73; this demonstrates the capacity of the vaccine candidates to induce specific antibody production that can recognize the native proteins.

To determine whether the selected proteins were effectively expressed and exposed on the surface of bacterial cells, a double immunofluorescence was carried out with the murine antisera and a specific antibody for *P. aeruginosa* anti-cell wall as initial characterization. Co-localization of the two signals could suggest that the proteins were present at the bacterial cell surface. As expected, sera of naïve mice did not recognize antigens while antisera of mice immunized with whole cell inactivated

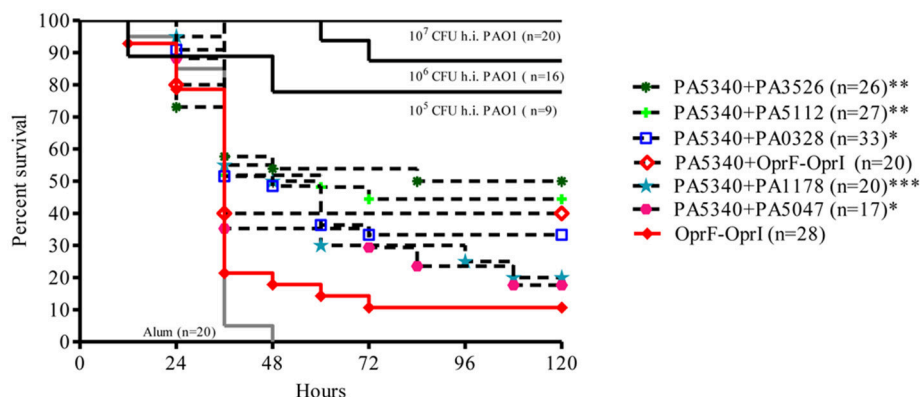


FIGURE 3 | Survival curves of groups of mice immunized with combined antigens selected as vaccine candidates. C57Bl/6 male mice were challenged with PAO1 (5×10^6 cfu) 2 weeks after last vaccination with combined antigens. Comparisons were performed with mice immunized with Alum alone (negative control) and PAO1 heat inactivated (h.i.) groups (positive control). An additional group was vaccinated with OprF-OprI, tested as clinical vaccine candidate. Data were pooled from at least two/three independent experiments ($n = 17-33$). Results are represented as Kaplan-Meier survival curves and analyzed by the Mantel-Cox test against negative control group: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. N refers to the number of animals.

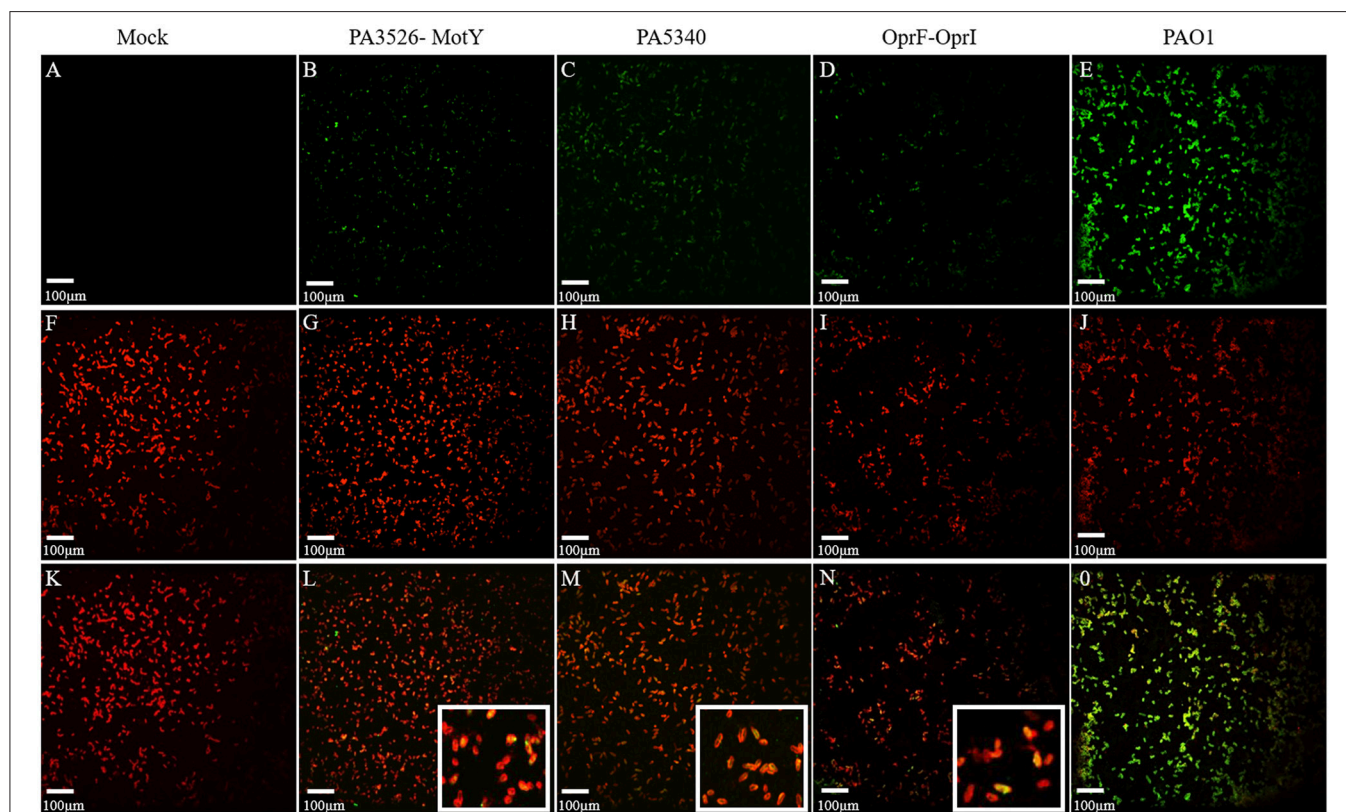


FIGURE 4 | Cellular localization of vaccine candidates PA5340 and PA3526-MotY and controls by immunofluorescence microscopy. Immunofluorescence staining with confocal microscopy shows the localization of antigens (green) (A–E) and the PAO1 cell wall (red) (F–J). For antigens localization the antisera of naïve mice (A) or immunized with PA3526-MotY (B), PA5340 (C), OprF-OprI (D) or heat inactivated PAO1 (E) were used. Merged images show the co-localization of the two signals (yellow) (K–O) suggesting that proteins could be surface exposed. Detailed co-localization of antigens of interest is shown in the magnification (L, M, N).

PAO1 showed a strong co-localization signal. The same staining was performed with the antisera of mice immunized with the 10 selected antigens and OprF-OprI (Table 1 and Figure 4).

According to immunofluorescence microscopy results, all the selected antigens co-localized with the cell surface antibody, with the exception of PA5112-EstA. In particular, a similar

co-localization pattern was observed with antisera of PA5340, PA3526-MotY, and OprF-OprI.

Conservation Profile of Selected Antigens Among a Collection of CF Clinical Isolates

Conservation of the selected antigens among *P. aeruginosa* genomes was initially considered as selection criteria and the candidates were checked in the ensemble of publicly available complete genome sequences. Moreover, we investigated the sequence conservation of the ten most protective antigens in a collection of 19 clinical strains isolated from CF patients at the onset of infection and after years of chronic colonization (Figure S3). Full-length sequences of the ten genes were obtained by PCR from most of the 19 strains. Nucleotide sequences were translated into the amino acid (aa) sequences and compared with PAO1 protein sequence. Overall, identity conservation levels were higher than 98% (Table 1 and Table S2), confirming a strong potential of these proteins to induce effective and cross-protective immunity among *P. aeruginosa* circulating strains.

DISCUSSION

As an alternative treatment to antibiotics, immunotherapy should represent an option to prevent MDR infections by *P. aeruginosa*. A universal protein-based vaccine against *P. aeruginosa* remains a critical unmet medical need. To identify possible antigens suitable for the development of a *P. aeruginosa* vaccine, we explored the reverse vaccinology approach integrated with additional genomic and bioinformatic approaches. This strategy involved an initial screening of target antigens on the basis of their putative cellular localization; this identified a large number of proteins (2,430 ORFs) predicted to be surface or membrane-associated to various extents. Note that the *P. aeruginosa* genome is larger than most sequenced bacteria (6.3 Mb) (5) and the resulting high number of 5,570 predicted protein encoding genes challenges this approach. Hence, additional criteria were included for antigen selection, with the aim of reducing the pool to a reasonable number of candidates and to rationalize the subsequent experimental steps. In particular, the absence of sequence similarity to proteins encoded by the commensal *E. coli* K12 strain, as well as sequence similarity to proteins involved in primary house-keeping and cell metabolism functions have been used as cutting edge to narrow the shortlist of antigens. Proteins displaying epitopes very similar to those present in widely-conserved proteins from human and mouse origin were also excluded as candidates, as they might be poorly immunogenic and have a high probability of behaving as self-antigens.

Considering that large sequence diversity characterizes *P. aeruginosa* genome, we tested the conservation of selected candidates among the *P. aeruginosa* complete genome sequences available on public databases. This analysis confirmed the presence and conservation of selected antigens in the core genome. Further genomic analysis considered sequence intraclonal diversity of selected antigens in clinical strains isolated from CF patients. It is recognized that *P. aeruginosa*

is an antigenically variable microorganism and can undergo phenotypic variation under changing environmental conditions, such as the airways in CF patients (23). In particular, the adaptation process generates unique lineages of *P. aeruginosa* pathogenic variants that differ systematically from environmentally-acquired strains and can escape immune recognition (24). We considered this question worthy of investigation and expanded the comparative sequence analysis on a collection of 19 clinical strains isolated from CF patients at the onset of infection and after years of chronic colonization (25, 26). The selected antigens were conserved among CF clinical isolates indicating that the corresponding genes were not under positive selection and these antigens could be useful for targeting both the initial infecting strains and those promoting progression toward chronic infection. Considering the epidemiology of *P. aeruginosa* infection, both environmental-to-host and patient-to-patient transmission have been described and it appears likely that highly conserved not-adapted antigens might have superior clinical relevance.

Previous studies that tested *P. aeruginosa* antigens provided valuable information on the feasibility of vaccination but were limited by either the number of antigens tested and by redundancy in their selection (10). Abundant surface proteins and oligosaccharides, particular virulence factors have been previously considered. However, none of these started from a large scale screening, performed comparative analysis of the *P. aeruginosa* protein antigens and tested in animal models for preclinical studies. Our screening identified 52 antigens distributed as having known (31) or unknown functions (21). The presence of known virulence factors in the final list of candidates confirmed the reliability of this general selection strategy. We identified a number of proteins, like ExoA and ExoT, as well as relevant outer membrane proteins, like OprF, and OprH, already shown to be required for virulence in *P. aeruginosa*. The outer membrane protein PscC precursor (PA1716) identified in this study was previously identified by integrated genomics and proteomics approach (11). Furthermore, different proteins involved in the chaperone/usher pathways (CupA-E) were identified in this study and Rashid et al. (11). Half of the candidate antigens identified belonged to the functional category of hypothetical, unknown, and unclassified genes (21 genes, 35.8% of the total), suggesting that there is still a large proportion of potentially immunogenic antigens to be discovered within the unexplored part of the *P. aeruginosa*.

Several diverse animal models have been used in preclinical studies of vaccination to evaluate *in vivo* protection against *P. aeruginosa* infection. Mouse models, including burned animals, those with immunocompromised intraperitoneal infection, or with acute pneumonia, have primarily been used in the past. These models were used for preclinical evaluation of candidates like the flagellum, the alginate exopolysaccharide conjugated to tetanus toxoid, polysaccharides, and outer membrane protein such as OprF and OprI. (27–31). Based on our previous experience we consider respiratory infection in immunocompetent mice a highly robust and appropriate model to predict efficacy of vaccine candidates for further clinical testing in patients at risk of respiratory infection,

such as VAP or CF (22, 32). The mouse model of acute infection has been extensively employed as the standard in *P. aeruginosa* pathogenesis and efficacy studies (24, 33–35). In this work, C57Bl/6 mice succumbed following infection with high dose of PAO1 virulent *P. aeruginosa* strain. Vaccination with whole cell inactivated *P. aeruginosa* induced an effective immunological response as demonstrated by full protection of the mice. Conversely, mice immunization with the adjuvant alone did not provide any protection as all mice showed severe clinical disease. These data strongly support the choice of this robust mouse model for screening and selecting the best vaccine candidates. Vaccination with the ten purified soluble proteins on our short list demonstrated distinct disease phenotypes, ranging from severe pneumonia to partial protection from a lethal dose of *P. aeruginosa* infection. We report that half of the vaccine candidates screened in this study were more effective when compared to OprF/OprI. It is worth noting that OprF/OprI fusion was one of the treatments evaluated in clinical trial although clinical efficacy has been disappointing (36). Given that vaccines containing several antigens have been shown to confer better protection than those containing only one antigen (20, 37), we decided to assess antigen combinations. Selection of antigens to combine was made from the shortlist of the ten most promising antigens and tested to further increase the vaccine efficacy and survival rate in murine models. This systematic screening identified five combinations, capable of significantly increasing the survival rate among challenged mice. All combinations included PA5340, a hypothetical protein exclusively present in *P. aeruginosa*. The maximum proportion of mice protected against challenge was 50%, achieved with PA5340 combined with PA3526-MotY. These two proteins were capable of triggering a specific immune response and initial characterization could indicate surface exposure. Nevertheless, as their function is still undetermined, it is unlikely that either protein would ever be selected by a traditional approach. Overall this study confirms the capability of reverse vaccinology to give new impetus in the research of vaccines against *P. aeruginosa* infection through the rapid identification of novel vaccine candidates.

MATERIALS AND METHODS

Ethics Statement

Animal studies adhered to the Italian Ministry of Health guidelines for the use and care of experimental animals (protocol #443). The use of the clinical data is in line with study no. 3739 that was approved by the Ethics Commission of Hannover Medical School.

In silico Analysis and Computational Tools

PSORTb (38) was used for the subcellular localization prediction, SignalP (39) to predict the SPs and their probable cleavage site in secreted proteins, the TatP prediction server to predict the presence of bacterial Tat signal peptides (40), LipoP (41) predict lipoproteins, TMPred to predict transmembrane segments (42).

To check presence and conservation of vaccine candidates in other *P. aeruginosa* genomes, comparative protein sequence

analysis against the sequenced genomes was performed using BLAST. The amino acid sequence of strain PAO1 was aligned against the protein translation of complete genome sequences with BLASTp (BLAST 2.2.26+) and against the genome of human and mouse with VAXGEN (<http://www.violinet.org/vaxgen/index.php>) (43).

Gene Sequencing

Reference strain PAO1 (5) and 19 *P. aeruginosa* clinical strains isolated from CF patients attending the Medizinische Hochschule Hannover and described previously (25, 26) were used to sequence the genes of vaccine candidates. PCR genes amplification was carried out using the list of primers detailed in the **Supplementary Information**.

Cloning, Expression, and Purification of *P. aeruginosa* Recombinant Proteins

Polypeptides antigens from *P. aeruginosa* PAO1 were PCR-amplified using specific oligonucleotides and *P. aeruginosa* chromosomal DNA as templates. Resulting PCR products were cloned in pET15b (Novagen) using the PIPE method (44). To express cloned proteins, BL21(DE3)T1^r clones containing pET15b constructs were grown in LB medium containing 100 µg/ml Ampicillin at 37°C until OD₆₀₀ = 0.5. Protein expression was induced by adding 1 mM IPTG and growing at the same temperature for additional 3 hrs. Conventional protein extractions and SDS-Page were performed to check protein expression. Western Blot was used to confirm proper expression of tested *P. aeruginosa* antigens.

Protein purification has been performed as reported previously (45). Briefly, bacteria cells undergone to mechanical or chemical lysis and recombinant polypeptides were recovered from crude cell extracts by immobilized-metal ion affinity chromatography (IMAC) using His MultiTrapTM HP 50 mL NiSephrose High-Performance 96 well-vacuum plates (GE Healthcare). Polypeptides expressed as insoluble inclusion bodies were solubilized in 50 mM Tris-HCl elution buffer, pH 8.8 containing 8 M Urea, 1 mM TCEP-HCl, and 250 mM imidazole. Renaturation was performed by dialysis in 50 mM NaH₂PO₄, pH 8.8 containing 10% (v/v) glycerol, 0.5 M arginine, 5.0 mM of reduced glutathione, 0.5 mM of oxidized glutathione in the presence of 4, 2, or 0 M urea. Protein concentration was determined using the Micro BCA protein assay reagent kit (Pierce). Protein purity was checked by SDS-PAGE CRITERION XT Precast Gel (Biorad) followed by Coomassie blue staining.

Mouse Immunizations and Protection Model

C57BL/6NCrIBR 5 week-old male mice (Charles River) were immunized i.p. at day 0, 21, and 35 with recombinant proteins formulated with Alum, either individually or as a combination of proteins. The formulations were optimized for pH and osmolarity. Each antigen was used at 10 µg/formulation/animal. The final concentration of Alum was 2 mg/ml in 10 mM histidine buffer (pH 6.5). Negative control mice were immunized with Alum alone, while positive control mice were boosted with

whole cell heat-inactivated PAO1 strain at different doses (10^5 – 10^7 CFU). To obtain the antisera, mice were bled at day -1, 34, and 49. At day 50, mice were challenged with 5×10^6 CFU of *P. aeruginosa* PAO1 strain by acute infection and monitored every 12 h for general wellness as detailed in the **Supplementary Information**.

Western Blot and Immunofluorescence Microscopy

Specific antisera from immunized mice were used to confirm protein expression by Western Blot and surface localization of antigens by immunofluorescence as detailed in the **Supplementary Information**.

Statistical Analysis

Statistical calculations and tests were performed using Mantel-Cox test and one-way ANOVA, considering $p < 0.05$ as the limit of statistical significance.

AUTHOR CONTRIBUTIONS

AB and VM: conceiving and designing the experiments. BA-F, IB, MS, SB, SM, AC, and MD: performing experiments. AB, BA-F, IB, MS, and VM: analyzing data and interpretation of the experiments results. AB, BA-F, IB, MS, and VM: writing the manuscript.

REFERENCES

- Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, et al. Bad bugs, no drugs: no ESKAPE! an update from the infectious diseases society of america. *Clin Infect Dis.* (2009) 48:1–12. doi: 10.1086/595011
- Williams BJ, Dehnhostel J, Blackwell TS. *Pseudomonas aeruginosa*: host defence in lung diseases. *Respirology* (2010) 15:1037–56. doi: 10.1111/j.1440-1843.2010.01819.x
- Gellatly SL, Hancock RE. *Pseudomonas aeruginosa*: new insights into pathogenesis and host defenses. *Pathog Dis.* (2013) 67:159–73. doi: 10.1111/2049-632X.12033
- Chastre J, Fagon JY. Ventilator-associated pneumonia. *Am J Respir Crit Care Med.* (2002) 165:867–903. doi: 10.1164/ajrccm.165.7.2105078
- Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warren P, Hickey MJ, et al. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* (2000) 406:959–64. doi: 10.1038/35023079
- Page MG, Heim J. Prospects for the next anti-*Pseudomonas* drug. *Curr Opin Pharmacol.* (2009) 9:558–65. doi: 10.1016/j.coph.2009.08.006
- Projan SJ, Youngman PJ. Antimicrobials: new solutions badly needed. *Curr Opin Microbiol.* (2002) 5:463–5. doi: 10.1016/S1369-5274(02)00364-8
- Holder IA. *Pseudomonas* immunotherapy: a historical overview. *Vaccine* (2004) 22:831–9. doi: 10.1016/j.vaccine.2003.11.028
- Sedlak-Weinstein E, Cripps AW, Kyd JM, Foxwell AR. *Pseudomonas aeruginosa*: the potential to immunise against infection. *Expert Opin Biol Ther.* (2005) 5:967–82. doi: 10.1517/14712598.5.7.967
- Döring G, Pier GB. Vaccines and immunotherapy against *Pseudomonas aeruginosa*. *Vaccine* (2008) 26:1011–24. doi: 10.1016/j.vaccine.2007.12.007
- Rashid MI, Naz A, Ali A, Andleeb S. Prediction of vaccine candidates against *Pseudomonas aeruginosa*: an integrated genomics and proteomics approach. *Genomics* (2017) 109:274–83. doi: 10.1016/j.ygeno.2017.05.001
- Sharma A, Krause A, Worgall S. Recent developments for *Pseudomonas* vaccines. *Hum Vaccin.* (2011) 7:999–1011. doi: 10.4161/hv.7.10.16369

FUNDING

This work was sponsored by Novartis Vaccines and Diagnostics Srl; in March 2015 the Novartis non-influenza Vaccines business was acquired by the GSK group of companies. The sponsor was involved in all stages of the study conduct and analysis. This work was supported in part by the Italian Cystic Fibrosis Research Foundation (FFC#08/2006 and FFC#10/2009) to AB with the contribution of the Delegazioni FFC of Como, Catania, Vittoria Ragusa, Latina and LIFC onlus Associazione regionale siciliana, in memory of Simone.

ACKNOWLEDGMENTS

The authors thank B. Tummeler (Klinische Forschergruppe, Medizinische Hochschule Hannover, Germany) for supplying the *P. aeruginosa* clinical strain and G. B. Pier for providing specific rabbit anti *P. aeruginosa* cell wall antibody (Brigham and Women's Hospital, Channing Labs, Boston, USA).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2018.03021/full#supplementary-material>

- Westritschnig K, Hochreiter R, Wallner G, Firbas C, Schwameis M, Jilma B, et al. A randomized, placebo-controlled phase I study assessing the safety and immunogenicity of a *Pseudomonas aeruginosa* hybrid outer membrane protein OprF/I vaccine (IC43) in healthy volunteers. *Hum Vacc Immunother.* (2014) 10:170–83. doi: 10.4161/hv.26565
- Döring G, Meisner C, Stern M. A double-blind randomized placebo-controlled phase III study of a *Pseudomonas aeruginosa* flagella vaccine in cystic fibrosis patients. *Proc Natl Acad Sci USA.* (2007) 104:11020–5. doi: 10.1073/pnas.0702403104
- Cryz SJ Jr, Lang A, Rüdeberg A, Wedgwood J, Que JU, Fürer E, et al. Immunization of cystic fibrosis patients with a *Pseudomonas aeruginosa* O-polysaccharide-toxin a conjugate vaccine. *Behring Inst Mitt.* (1997) 98:345–9.
- Pier GB, DesJardin D, Grout M, Garner C, Bennett SE, Pekoe G, et al. Human immune response to *Pseudomonas aeruginosa* mucoid exopolysaccharide (alginate) vaccine. *Infect Immunology* (1994) 62:3972–9.
- Mora M, Veggi D, Santini L, Pizza M, Rappuoli R. Reverse vaccinology. *Drug Discov Today* (2003) 8:459–64. doi: 10.1016/S1359-6446(03)02689-8
- Scarselli M, Giuliani MM, Adu-Bobie J, Pizza M, Rappuoli R. The impact of genomics on vaccine design. *Trends Biotechnol.* (2005) 23:84–91. doi: 10.1016/j.tibtech.2004.12.008
- Grifantini R, Bartolini E, Muzzi A, Draghi M, Frigimelica E, Berger J, et al. Previously unrecognized vaccine candidates against group B meningococcus identified by DNA microarrays. *Nat Biotechnol.* (2002) 20:914–20. doi: 10.1038/nbt728
- Giuliani MM, Adu-Bobie J, Comanducci M, Aricò B, Savino S, Santini L, et al. A universal vaccine for serogroup B meningococcus. *Proc Natl Acad Sci USA* (2006) 103:10834–9. doi: 10.1073/pnas.0603940103
- Gabelsberger J, Knapp B, Bauersachs S, Enz UI, von Specht BU, Domdey H, et al. A hybrid outer membrane protein antigen for vaccination against *Pseudomonas aeruginosa*. *Behring Inst Mitt.* (1997) 98:302–14.
- Bragonzi A. Murine models of acute and chronic lung infection with cystic fibrosis pathogens. *IJMM* (2010) 300:584–93. doi: 10.1016/j.ijmm.2010.08.012

23. Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR, D'Argenio DA, et al. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc Natl Acad Sci USA* (2006) 103:8487–92. doi: 10.1073/pnas.0602138103
24. Cigana C, Lorè NI, Riva C, De Fino I, Spagnuolo L, Sipione B, et al., Tracking the immunopathological response to *Pseudomonas aeruginosa* during respiratory infections. *Sci Rep.* (2016) 6:21465. doi: 10.1038/srep21465
25. Bragonzi A, Wiehlmann L, Klockgether J, Cramer N, Worlitzsch D, Döring G, et al. Sequence diversity of the mucABD locus in *Pseudomonas aeruginosa* isolates from patients with cystic fibrosis. *Microbiology* (2006) 152:3261–9. doi: 10.1099/mic.0.29175-0
26. Bragonzi A, Wiehlmann L, Klockgether J, Cramer N, Worlitzsch D, Döring G, et al. *Pseudomonas aeruginosa* microevolution during cystic fibrosis lung infection establishes clones with adapted virulence. *Am J Res Crit Care Med.* (2009) 180:138–45. doi: 10.1164/rccm.200812-1943OC
27. Zuercher AW, Horn MP, Wu H, Song Z, Bundgaard CJ, Johansen HK, et al. Intranasal immunisation with conjugate vaccine protects mice from systemic and respiratory tract infection with *Pseudomonas aeruginosa*. *Vaccine* (2006) 24:4333–42. doi: 10.1016/j.vaccine.2006.03.007
28. Behrouz B, Hashemi FB, Fatemi MJ, Naghavi S, Irajian G, Halabian R, et al. Immunization with bivalent flagellin protects mice against fatal *Pseudomonas aeruginosa* pneumonia. *J Immunol Res.* (2017) 2017:5689709. doi: 10.1155/2017/5689709
29. Kashef N, Behzadian-Nejad Q, Najari-Peerayeh S, Mousavi-Hosseini K, Moazzeni M, Djavid GE, et al. Synthesis and characterization of *Pseudomonas aeruginosa* alginate-tetanus toxoid conjugate. *J Med Microbiol.* (2006) 55(Pt 10):1441–6. doi: 10.1099/jmm.0.46696-0
30. von Specht BU, Knapp B, Muth G, Bröker M, Hungerer KD, Diehl KD, et al. Protection of immunocompromised mice against lethal infection with *Pseudomonas aeruginosa* by active or passive immunization with recombinant P. aeruginosa outer membrane protein F and outer membrane protein I fusion proteins. *Infect Immun.* (1995) 63:1855–62.
31. Scarff JM, Goldberg JB. Vaccination against *Pseudomonas aeruginosa* pneumonia in immunocompromised mice. *Clin Vacc Immunol.* (2008) 15:367–75. doi: 10.1128/CI.00419-07
32. Kukavica-Ibrulj I, Facchini M, Cigana C, Levesque RC, Bragonzi A. Assessing *Pseudomonas aeruginosa* virulence and the host response using murine models of acute and chronic lung infection. *Methods Molecul Biol.* (2014) 1149:757–71. doi: 10.1007/978-1-4939-0473-0_58
33. Lore NI, Cigana C, De Fino I, Riva C, Juhas M, Schwager S, et al. Cystic fibrosis-niche adaptation of *Pseudomonas aeruginosa* reduces virulence in multiple infection hosts. *PLoS ONE* (2012) 7:e35648. doi: 10.1371/journal.pone.0035648
34. De Simone M, Spagnuolo L, Lorè NI, Rossi G, Cigana C, De Fino I, et al. Host genetic background influences the response to the opportunistic *Pseudomonas aeruginosa* infection altering cell-mediated immunity and bacterial replication. *PLoS ONE* (2014) 9:e106873. doi: 10.1371/journal.pone.0106873
35. Cigana C, Bernardini F, Facchini M, Alcalá-Franco B, Riva C, De Fino I, et al. Efficacy of the novel antibiotic POL7001 in preclinical models of *Pseudomonas aeruginosa* Pneumonia. *Antimicrob Chemother.* (2016) 60:4991–5000. doi: 10.1128/AAC.00390-16
36. Rello J, Krenn CG, Locker G, Pilger E, Madl C, Balica L, et al. A randomized placebo-controlled phase II study of a *Pseudomonas* vaccine in ventilated ICU patients. *Crit Care* (2017) 21:22. doi: 10.1186/s13054-017-1601-9
37. Olin P, Rasmussen F, Gustafsson L, Hallander HO, Heijbel H. Randomised controlled trial of two-component, three-component, and five-component acellular pertussis vaccines compared with whole-cell pertussis vaccine. Ad Hoc group for the study of pertussis vaccines. *Lancet* (1997) 350:1569–77. doi: 10.1016/S0140-6736(97)06508-2
38. Gardy JL, Brinkman FS. Methods for predicting bacterial protein subcellular localization. *Nat Rev Microbiol.* (2006) 4:741–51. doi: 10.1038/nrmicro1494
39. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* (2011) 8:785–6. doi: 10.1038/nmeth.1701
40. Bendtsen JD, Nielsen H, Widdick D, Palmer T, Brunak S. Prediction of twin-arginine signal peptides. *BMC Bioinform.* (2005) 6:167. doi: 10.1186/1471-2105-6-167
41. Juncker AS WH, Von Heijne G, Brunak S, Nielsen H, Krogh A. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* (2003) 12:1652–62. doi: 10.1110/ps.0303703
42. Hofmann KSW. TMbase - a database of membrane spanning proteins segments. *Biol Chem. Hoppe-Seyler* (1993) 374:166.
43. He Y, Racz R, Sayers S, Lin Y, Todd T, Hur J, et al. Updates on the web-based VIOLIN vaccine database and analysis system. *Nucleic Acids Res.* (2014) 42:D1124–32. doi: 10.1093/nar/gkt1133
44. Bonacci S, Buccato S, Maione D, Petracca R. Successful completion of a semi-automated enzyme-free cloning method. *J Struct Funct Genom.* (2016) 17:57–66. doi: 10.1007/s10969-016-9207-z
45. Palumbo EI, Fiaschi L, Brunelli B, Marchi S, Savino S, Pizza M. Antigen identification starting from the genome: a “Reverse Vaccinology” approach applied to MenB. *Methods Mol Biol.* (2012) 799:361–403. doi: 10.1007/978-1-61779-346-2_21

Conflict of Interest Statement: MS, SM, SB, AC, and VM were employees of Novartis Vaccines and Diagnostics Srl at the time of the study (now part of the GSK group of companies). They are now employees of the GSK group of companies. MD is a student at the University of Torino and participated in a post graduate studentship program at GSK.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Bianconi, Alcalá-Franco, Scarselli, Dalsass, Buccato, Colaprico, Marchi, Masignani and Bragonzi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Impact of Nucleotide Sequence Analysis on Meningococcal Vaccine Development and Assessment

Martin Christopher James Maiden*

Department of Zoology, University of Oxford, Oxford, United Kingdom

OPEN ACCESS

Edited by:

Pedro A. Reche,
Complutense University of Madrid,
Spain

Reviewed by:

Scott D. Gray-Owen,
University of Toronto, Canada
William William Shafer,
Emory University School of Medicine,
United States

*Correspondence:

Martin Christopher James Maiden
martin.maiden@zoo.ox.ac.uk

Specialty section:

This article was submitted to
Vaccines and Molecular Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 19 October 2018

Accepted: 20 December 2018

Published: 15 January 2019

Citation:

Maiden MCJ (2019) The Impact of
Nucleotide Sequence Analysis on
Meningococcal Vaccine Development
and Assessment.
Front. Immunol. 9:3151.
doi: 10.3389/fimmu.2018.03151

Since it became available as a routine tool in biology, the determination and analysis of nucleotide sequences has been applied to the design of vaccines and the investigation of their effectiveness. As vaccination is primarily concerned with the interaction of biological molecules with the immune system, the utility of sequence data is not immediately obvious and, indeed, nucleotide sequence data are most effective when used to complement more conventional immunological approaches. Here, the impact of sequencing on the field of vaccinology will be illustrated with reference to the development and implementation of vaccines against *Neisseria meningitidis* (the meningococcus) over the 30-year period from the late-1980s to the late-2010s. Nucleotide sequence-based studies have been important in the fight against this aggressive pathogen largely because of its high genetic and antigenic diversity, properties that were only fully appreciated because of sequence-based studies. Five aspects will be considered, the use of sequence data to: (i) discover vaccine antigens; (ii) assess the diversity and distribution of vaccine antigens; (iii) determine the evolutionary and population biology of the organism and their implications for immunization; and (iv) develop molecular approaches to investigate pre- and post-vaccine pathogen populations to assess vaccine impact. One of the great advantages of nucleotide sequence data has been its scalability, which has meant that increasingly large data sets have been available, which has proved invaluable in the investigation of an organism as diverse and enigmatic as the meningococcus.

Keywords: *Neisseria meningitidis*, conjugate polysaccharide vaccines, outer membrane vesicle vaccines, population biology, herd immunity, efficacy

INTRODUCTION

The 40 years following the introduction of the Sanger dideoxy method in 1977 (1) saw a revolution in biology, which was driven by the improvement of nucleotide sequencing technologies. At the start of this period, determining a DNA or RNA sequence was a highly specialized task, which was undertaken in a very few laboratories most often using their own home-made equipment and reagents. Only individual genes or viruses could be sequenced, and then at great expense. By 2018, nucleotide sequencing was conducted on an industrial scale, employing mass-produced reagents and highly automated equipment, often in large factory-like installations. Complete genome sequences were assembled routinely for tens of thousands, even hundreds of thousands of organisms, including those with the largest genomes. Major advances had also been made in the

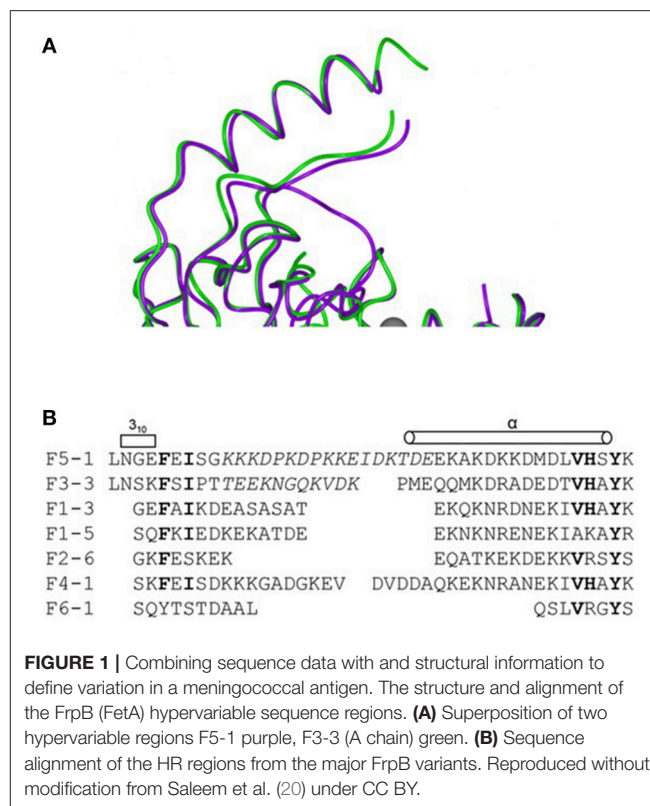
computer equipment and software available to interpret the data produced, although it is fair to say that while issues of data generation were for all practical purposes resolved, data interpretation remained a major hurdle. In common with most areas of biology, vaccine development, and evaluation were transformed by these advances and this transformation will be illustrated here using meningococcal vaccines as an example.

Many of the most successful bacterial and viral vaccines were developed in the mid- to late- twentieth century, without recourse to the detailed genetic information that nucleotide sequencing provides; however, most of these conventionally-developed vaccines targeted antigenically stable pathogens, such as the smallpox virus, or those that rely on a single, stable, molecule for their pathogenicity, such as *Corynebacterium diphtheriae* (diphtheria) and *Clostridium tetani* (tetanus). The bacterium *Neisseria meningitidis*, the meningococcus, very nearly falls in to this category, as almost all invasive meningococcal disease (IMD) is caused by bacteria that express one of only six capsular polysaccharide antigens, referred to as serogroups A, B, C, W, X, and Y. Polysaccharide and, especially, protein-conjugate polysaccharide, vaccines are effective in protecting against disease for five of these (serogroups A, C, X, W, X, and Y) and provide a means of eliminating most IMD worldwide (2). Unfortunately, however, vaccines directed against meningococcal serogroup B polysaccharide have not been developed. This is a consequence of poor immunogenicity and fear of inducing host autoimmune disease, due to similarities of the serogroup B polysaccharide with human antigens (3, 4). The search for alternative antigens that target serogroup B meningococci has been complicated by the high variability of virtually all other possible meningococcal vaccine components. Ever-increasing volumes of nucleotide sequence data have been used in this search.

As successful immunization is almost always a population process, population studies based on nucleotide sequences have important applications in the development, deployment, and evaluation of vaccines. Here I shall describe how nucleotide sequence technologies have contributed to vaccine development and the assessment of vaccines, outlining the development, and implementation of meningococcal vaccines from the late 1970s to the time of writing.

STRUCTURE AND VARIATION IN INDIVIDUAL ANTIGENS

Determining the nucleotide sequence of a gene encoding a vaccine antigen permits the deduction of a wealth of information concerning the protein and enables a wide range of follow-up studies. For example, the PorA protein was established as a potential vaccine component for “serogroup B substitute” vaccines (i.e., vaccines developed as an alternative to those including serogroup B capsular polysaccharide as an antigen) in the late 1970s and early 1980s, on the basis of immunological and biochemical studies (5, 6). The cloning and sequencing of the *porA* gene in 1991 provided much additional information, for example confirming that it was a typical porin related to those



found in the gonococcus (7). The meningococcal and gonococcal sequences were used to design oligonucleotide primers for the then new PCR technique, enabling the amplification, and sequencing of the *porA* gene from multiple meningococcal isolates (8), illustrating how sequence technologies lead to cumulative gains in knowledge. These comparative studies enabled structural models to be proposed and established that the antigenic variability identified by subtype-specific monoclonal antibodies was mostly determined by the peptide sequences of two major variable regions (VR1 and VR2) and one less variable region (VR3 or sVR) of the PorA protein. Combining sequencing and immunological studies enabled these antigens and their interaction with immune molecules to be defined (8–10).

The combination of PCR amplification and Sanger sequencing enables multiple variants of a given gene to be characterized accurately and rapidly at high volume. This permits the antigenic variability of a given protein vaccine component to be established. In the case of meningococcal PorA, the variants of which were initially identified with monoclonal antibodies (11), an ever-increasing diversity of variants have been identified by sequencing in the past 30 years. This required that the original nomenclature, which was based on monoclonal antibody reactivity, needed to be replaced with an updated nomenclature scheme that was based on the peptide sequences themselves, rather than the antibodies that reacted with them (12). It was important that this scheme was backwards compatible with the designations of the established antibody-based system. The scheme needed to be infinitely expandable, and to this end

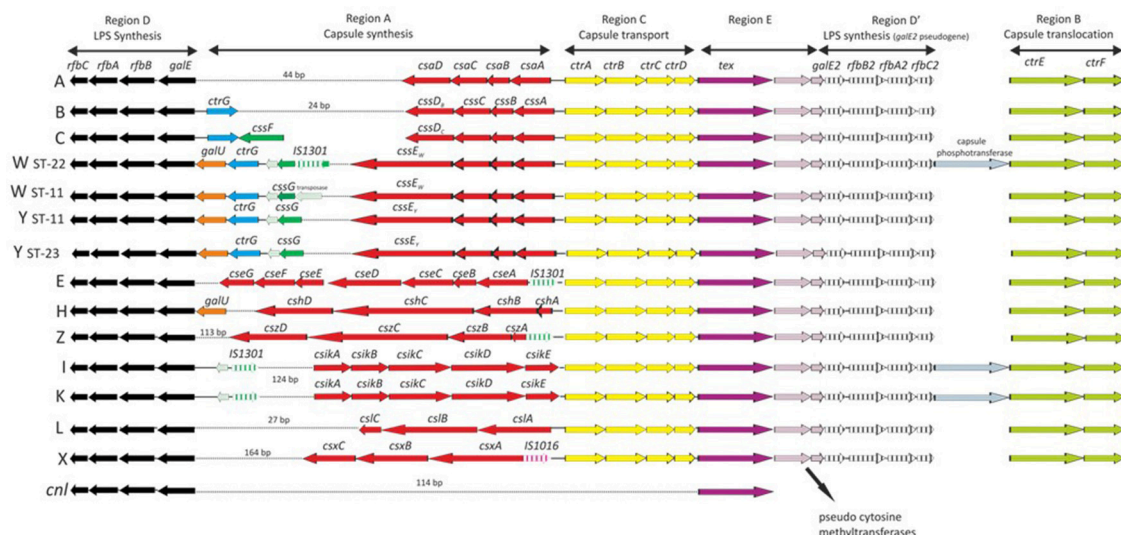


FIGURE 2 | Genetic organization of the *cps* locus among *Neisseria meningitidis* serogroups revealed by nucleotide sequencing. Serogroups A (*N. meningitidis* isolate Z2491); B (H44/76); C (FAM18, 053442, and 29013); W (α 275); W (WUE171); Y (α 162); Y (WUE172); E (α 707); H (29031); I (29043); K (29046); L (WUE3608); X (α 388); and Z (WUE173); and a capsule null (*cni*) isolate. Letters on left represent serogroups. Arrows depict gene orientation. Reproduced without modification from Harrison et al. (43) under CC BY.

variant numbers were used, which could be added to as each new variant was discovered. At the time of writing at total of 324 peptide variants had been defined or PorA VR1 and 906 for PorA VR2. An advantage of sequence-based nomenclature is that the sequences can be recorded on open-access web based databases (e.g., <https://pubmlst.org/neisseria/PorA/>), enabling easy access to the nomenclature (13). Similar approaches have since been used to catalog the variation of a number of meningococcal vaccine candidates including: factor H binding protein (fHbp) (14); the ferric enterochelin receptor, FetA (15); *Neisseria* adhesin A (NadA) (16); and the heamoglobin receptor (HmbR) (17).

In addition to describing the nature and extent of variation of genes, including those encoding vaccine antigens, sequence analyses help to reveal the mechanisms whereby this variation occurs. In the case of PorA, the immunogenic VRs are relatively short continuous sequences, encoding surface-exposed parts of the porin structure (18), which vary by point mutation, insertion, and deletion. Each of these processes can have an impact on the binding of immune molecules such as antibodies to the expressed protein. These impacts have been assessed by a combination of sequence comparison, biochemical, structural, and immunological analyses (19, 20) (Figure 1). Sequencing studies have also established that protein antigen expression can be influenced by the sequence of control regions, with polynucleotide tracts playing an important role in the expression of a number of meningococcal antigens (21). Another mechanism of variation that sequencing studies identified is the exchange of genetic material *via* horizontal genetic transfer (HGT), and the PorA protein was one of the first bacterial genes in which this was extensively documented (22). The recognition of the importance of HGT in bacterial evolution came from studies of antibiotic resistance and vaccine antigens

in the gonococcus (23), meningococcus (22), and pneumococcus (24), with a major impact on our understanding on bacterial population biology and evolution (25).

BACTERIAL POPULATION BIOLOGY AND EVOLUTIONARY STUDIES

Knowledge of the extent of variation in vaccine antigens inevitably poses questions of how this variation comes about and how it moves through the population. Both topics have important implications for vaccine design, as it is essential to know how effective a given vaccine will be once it has been introduced and how easily vaccine escape variants might arise and spread. Population studies of the pathogenic *Neisseria*, both the meningococcus and *Neisseria gonorrhoeae* (the gonococcus), have played a major role in developing bacterial population genetics. In the case of the meningococcus, these studies have also been central in the design and implementation of new vaccines.

Bacterial population studies predate the sequencing era, with seminal investigations of the meningococcus using a combination of multi-locus enzyme electrophoresis (MLEE) (26) and immunological typing (27–29), establishing fundamental concepts that have been built on subsequently by sequence-based investigations. MLEE investigations conducted on isolates from cases of invasive disease, showed that a limited number of groups of closely related organisms, known as “clones” or “genetic lineages,” each of which was associated with particular antigenic characteristics, were predominant causes of disease (27, 29–32). By contrast, meningococci isolated from asymptomatic carriers were much more diverse (33, 34). This led to the concept of “hypervirulent” or “hyperinvasive” meningococci: persistent

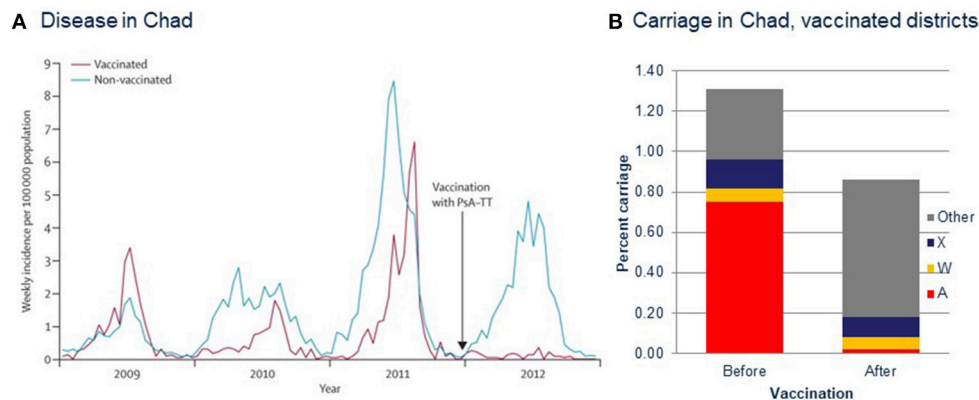


FIGURE 3 | Measuring vaccine efficacy in the meningitis belt during PsA-TT (MenAfriVac) introduction in Chad using genomic sequencing techniques. **(A)** Incidence of reported cases of meningitis in Chad, 2009–2012. Vaccination with PsA-TT was undertaken in patients aged 1–29 years at the end of 2011 (arrow). PsA-TT=serogroup A meningococcal polysaccharide–tetanus toxoid conjugate vaccine. **(B)** Carriage rates of meningococci of different serogroups before and after vaccination. **(A)** Reproduced from Daugla et al. (68) without modification under CC BY.

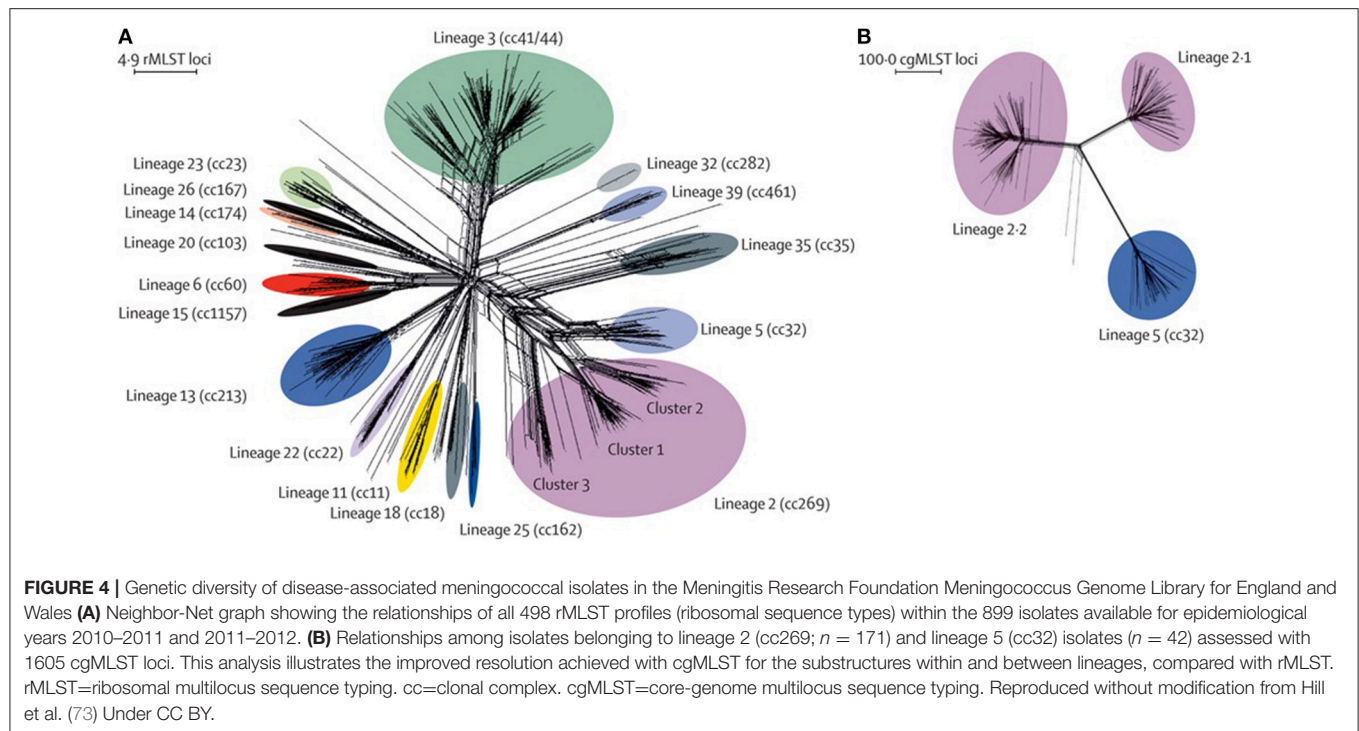
genotypes that were especially likely to cause IMD (35). One such lineage, associated with electrophoretic type 5 (ET-5), caused an international outbreak of serogroup B IMD (30), resulting in the development and implementation of specific outer membrane vesicle (OMV) vaccines generated against the local outbreak strain in Norway (36) and Cuba (37). This approach was repeated in New Zealand 15 years later (38).

The high levels of HGT observed in genes encoding antigens (22, 39) and antibiotic resistance (23), sometimes involving inter-species transfer (23, 40), are also observed in “housekeeping” genes: those expected to be subject to stabilizing selection for the conservation of metabolic function (41). This demonstrates the importance of HGT in bacterial evolution generally, and especially in the *Neisseria*. One consequence of this is the a range of population structures observed in the meningococcus (25). MLEE technology was difficult to implement and compare across studies and, as sequencing technology developed, it was replaced by a nucleotide sequence-based approach, multilocus sequence typing (MLST), which indexed the variation in seven ~400 bp fragments of housekeeping genes: *abcZ*; *adh*; *aroE*; *fumC*; *gdh*; *pdhC*; and *pgm*. MLST was much easier to implement at scale and had the advantage of being easily portable among laboratories (42). Originally developed for *Neisseria meningitis*, MLST has had wide application to many bacterial species and, in combination with web-accessible databases, has become a widely used method, with applications in, for example: evolutionary studies; epidemiology; population biology; and taxonomy (42). MLST studies enabled genetic lineages to be defined in terms of groups of related STs called clonal complexes (ccs), which were named after the ST that was used to define that complex. Thus, ST-11 is the defining sequence type for clonal complex 11 (cc11), also known as the “ST-11 complex.” These molecular designations have been incorporated with other typing schemes such as serogrouping (for capsule) (43) (Figure 2), subtyping (for PorA) (12) and typing of the FetA antigen (Figure 1) (15) into a standardized typing nomenclature with the form:

B: P1.19,15; F5-1; ST-33 (cc32), referring to: group (in this case B); PorA VR1 and VR2 type (P1.19,15); FetA VR type (F5-1); sequence type (ST-33); and clonal complex (cc32), respectively (44).

THE POPULATION APPROACH TO VACCINE ASSESSMENT

High-throughput sequencing approaches, which permit the characterization of hundreds or thousands of bacteria isolates at multiple loci, enable a population genetic approach to be taken to the assessment of vaccination programs. In the case of *N. meningitidis*, these approaches played a central role in understanding the impact of conjugate polysaccharide vaccines, which targeted the meningococcal capsular antigen (2). In late 1999, just after the publication of the MLST approach in 1998, the United Kingdom (UK) Department of Health implemented a novel monovalent serogroup C conjugate polysaccharide vaccine, in response to a dramatic increase in serogroup C meningococcal disease in the UK (45). The protein-polysaccharide vaccines had significant advantages over the older “plain” polysaccharide vaccines, in that the conjugate vaccines generated more effective immune responses. This included eliciting protective responses in younger individuals, affinity maturation, and a memory response, none of which were generated by the “plain” polysaccharide vaccines (46). A combination of the urgency of the public health response and the epidemiology of IMD meant that the new meningococcal C conjugate (MCC) vaccines were implemented without the benefit of a phase III efficacy trial (2). This led to several uncertainties, not least that it was unknown whether the vaccine would affect asymptomatic carriage. Although the prospect of directly preventing disease was wholly positive, there was a concern that population effects might lead to the emergence of vaccine escape variants—variants of the epidemic clone that had acquired a different capsule by



HGT (47). Investigation of this required monitoring, not only disease isolates to establish impact on disease, but also carried meningococci, during vaccine implementation period.

The United Kingdom Meningococcal Carriage (UKMenCar) study was established to measure the impact of MCC on carriage and surveyed 47,765 teenagers from 1999 to 2001, immediately before and for 2 years after vaccine implementation (48–50). Oropharyngeal swab samples and conventional culturing were used to isolate meningococci from the throats of the subjects. In combination with phenotyping, high-throughput single-locus sequencing of the capsular region combined with MLST analysis enabled the nature and status of the serogroups to be determined, together with the clonal complex of each isolate. These data showed that, before vaccine implementation, the point prevalence of carriage of the epidemic-causing meningococcus (C:cc11) was low (0.31% of individuals), which is characteristic of a highly hyperinvasive meningococcus. In 2001, 2 years after implementation, this carriage rate decreased by 80% (to 0.04% of individuals). As a proportion of the meningococci isolated, the C:cc11 epidemic strain dropped by nearly 90%, from 1.83 to 0.21%, with a coincident reduction in the proportion expressing the capsular antigen of around 50% (48). This resulted in a significant herd immunity (protection) of the unvaccinated (51), which was fortunate as the initial vaccination schedule used for infants (2, 3, and 4 months) provided no direct protection more than a year after immunization (52). The vaccination program had no significant effect on the carriage of other genotypes and serogroups, although there was some evidence for secular changes over the period of the surveys (50).

The success of the monovalent MCC vaccines, demonstrated to be in large degree due to their ability to induce herd immunity, catalyzed interest in developing a similar vaccine to target the periodic very large epidemics seen in the African Meningitis Belt. These represent one of the most important manifestations of IMD globally and were first described by Lapeyssonnie (53). Since that time, repeated serogroup A epidemics had been observed, with especially large outbreaks in the late 1990s (54). The meningitis vaccine project (MVP) was established in 2001 as a partnership between PATH (formerly the Program for Appropriate Technology in Health) and the World Health Organization (WHO), funded by the Bill and Melinda Gates foundation. The aim of the MVP was the elimination of meningococcal epidemics in Africa by means of an affordable serogroup A protein-polysaccharide conjugate vaccine. Employing an innovative public-private approach with northern and southern partners, this goal was achieved, with the vaccine PsA-TT, “MenAfriVac®,” introduced in December 2010, with a plan to immunize all African Meningitis Belt countries by 2013 (55, 56).

The Meningococcal Carriage in Africa (MenAfriCar) consortium was established in 2008 to measure the impact of the PsA-TT vaccine on carriage and the impact of herd immunity (57). Up until that time, knowledge of the carriage of meningococci in this region was incomplete, with a number of studies undertaken at various times employing a variety of techniques. This limited the collation of consistent information and a wide range of carriage prevalence rates (3–30%) had been reported (58). The MenAfriCar consortium aimed to conduct large-scale carriage surveys across the meningitis belt before and

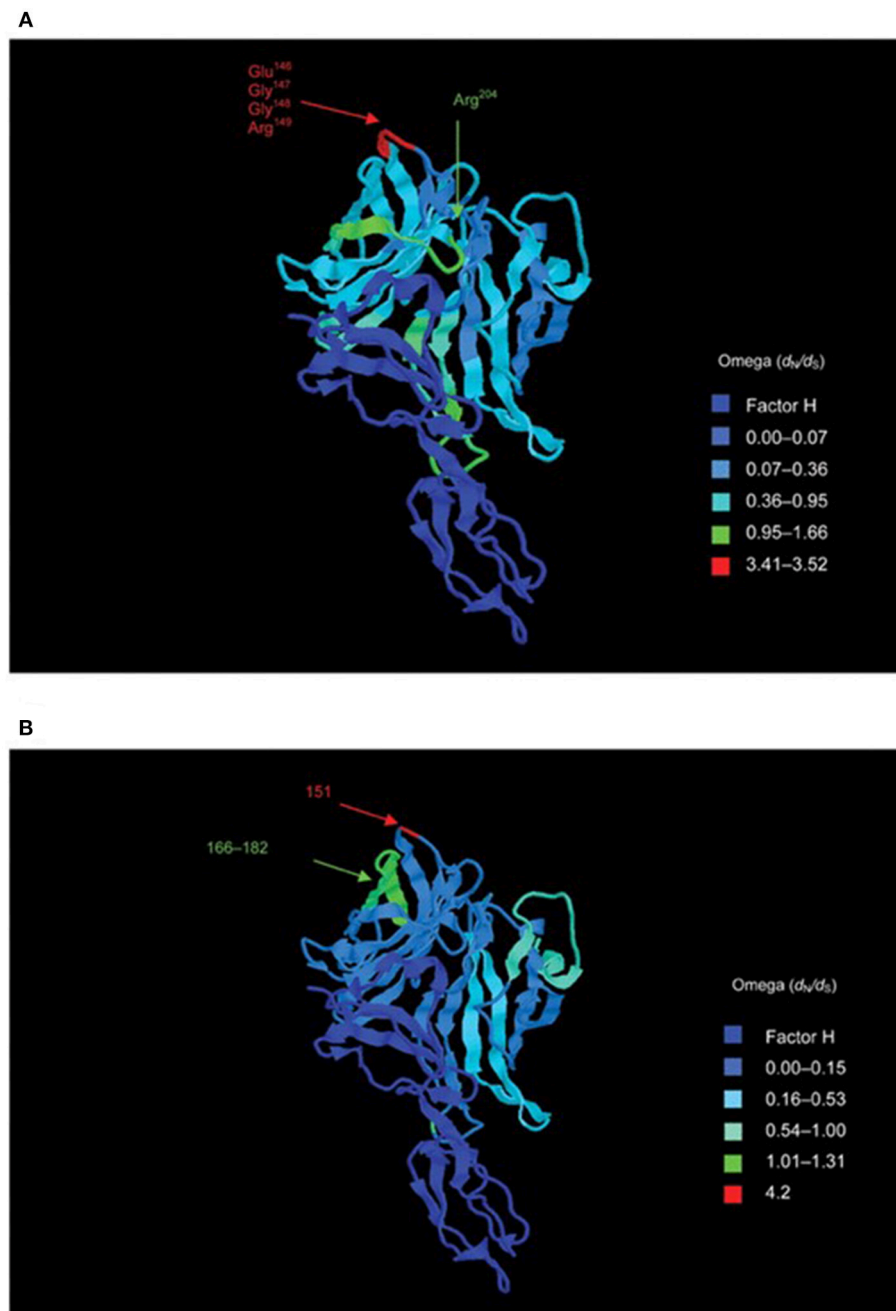


FIGURE 5 | Structure of the fH-fHbp complex, with temperature coloring using per-site point estimates of selection (ω) for **(A)** subfamily B/variant 1 sequences and **(B)** subfamily A/variant 2 sequences. Peptides indicated in **(A)** are putative bactericidal epitopes. In **(B)** positively selected sites are indicated. Note: subfamily A/variant 2 differs in length from subfamily B/variant 1 by +4 bp (e.g., Glu151 is equivalent to Glu147 in variant 1). Reproduced from Brehony et al. (14) without modification under CC BY.

after the introduction of the PsA-TT vaccine using consistent methods. Risk factor data for carriage were simultaneously collected, as in the UKMenCar surveys, although the risk factors included were somewhat different from those seen in the UK (59, 60). One unknown was the impact of the carriage of other *Neisseria* species, especially *Neisseria lactamica*, which

is commonly isolated from individuals in Africa (61). This was challenging as members of the genus are very similar: they are poorly distinguishable by 16SrRNA sequencing (62), for example. The large number of isolates that would have to be processed by laboratories in resource-limiting environments presented further challenges that were met by the exploitation

of nucleotide sequence-based approaches, made possible by the availability of genomic technologies.

The MenAfriCar surveys employed a combination of conventional culture, biochemical, and sequence-based methodologies. As in UKMenCar surveys, oropharyngeal swab samples were collected and cultured on selective media with putative *Neisseria* identified using biochemical tests in local laboratories (60). From these cultures, boiled cell preparations were made, which were shipped to Oxford where the molecular analyses were performed (61). To solve the speciation problem a novel assay was designed. This took advantage of an extended MLST scheme, ribosomal MLST (rMLST), which indexes the 53 genes encoding ribosomal proteins, and enables bacterial isolate characterization “from domain to strain” (63). From the rMLST sequences extracted from WGS data from 44 isolates of diverse species, a 413 bp fragment of the *rplF* gene was identified, the sequence of which was diagnostic for each *Neisseria* species. This fragment could be amplified and sequenced at high-throughput, enabling *Neisseria* species identification which was rapid, accurate, and cost-effective (64). The capsule genes were identified with a real time PCR assay and fine typing performed by sequencing the *porA* and *fetA* loci (60).

The MenAfriCar studies showed great diversity of meningococcal carriage across the belt and over time, which differed from the patterns of carriage typically observed in high-income countries with temperate climates, where carriage is more consistent (65). The age profiles of carriers was also different, with meningococcal carriage highest in individuals aged 5–14 (60), rather than in adults and adolescents as typically seen elsewhere (65). Transmission among children within households was shown to be important (66), again different from other settings, where social interactions outside the family are important (59). There was also much more diversity in the non-meningococcal *Neisseria* species isolated, the carriage of which was also age-related (61). This different dynamic was consistent with the unique epidemiology observed in African meningitis belt countries, and with the idea that the occurrence of seasonal epidemics in the meningitis belt were dependent on the transmission of epidemic clones.

Most of the counties in which PsA-TT was introduced including the first, Burkina Fasso, were not experiencing an epidemic of serogroup A IMD at the time of introduction, making assessment of the herd effects of the vaccine difficult, although effects on disease and carriage rates were consistent with such effects (67). An epidemic in Chad during the introduction there, however, enabled a direct demonstration of efficacy against IMD and carriage of the epidemic strain (68) (Figure 3). The rollout of the vaccines in different districts in over two epidemic years, combined with the MenAfriCar sampling and sequence-based isolate characterization protocols, demonstrated high efficacy of the vaccination against both IMD and carriage (68). Samples collected in this study also demonstrated that even at genomic levels of isolate characterization, there were no consistent differences between carried and invasive meningococci (69).

The marked successes of the conjugate vaccines in different settings provided the prospect of a “meningitis free world,”

so long as an effective group B vaccine could be developed (70); however, no vaccine against the meningococcal serogroup B polysaccharide, either plain or protein-conjugate, has been developed to the time of writing and none was under development (3). This was due to a combination of the poor immunogenicity of the antigen combined with fears of the induction of auto-immunity, as a consequence of its similarity of the antigen to human neural polysaccharides (4). Consequently, even with the prospect of conjugate polysaccharide vaccines that target serogroups A, C, W, X, and Y (71), there is a need for alternative or “substitute serogroup B” vaccines, if IMD is to be comprehensively prevented (3).

GENOMIC DISEASE SURVEILLANCE: UNDERSTANDING AND COMBATTING VIRULENCE AND VACCINE ESCAPE

The development and validation of next generation sequencing approaches for the determination of high-quality draft WGSs of meningococci (72), led to the establishment of the Meningitis Research Foundation Meningococcus Genome Library (MRF-MGL) (73) (Figure 4). This repository contained the WGS data for all meningococci isolated from cases of IMD in the UK. The MRF-MGL provides the opportunity to identify and react to IMD outbreaks in real time or near real time. An example of this is the reaction of the UK public health authorities to a serogroup W IMD outbreak. From the early 2010s onwards, coinciding with the establishment of the MRF-MGL, there was a year-on-year increase in cases of serogroup W meningococcal disease, which data from the MRF-MGL showed to be W:cc11 (74). A similar increase had been observed a decade before (75), just after the successful introduction of the MCC vaccines, causing some alarm (76). The former increase had been associated with the global spread of a particular W:cc11 strain after the Hajj pilgrimage and had fortunately been transitory (77); however, the epidemiology of the cases in the 2010s was somewhat different, leading to the concern that this might be a different circumstance and that a larger epidemic might occur (74). A WGS study of a global collection of 750 diverse cc11 meningococci demonstrated that, although the cases in the UK were indeed related to the Hajj-associated isolates, they were much more closely related at the WGS level to W:cc11 meningococci that had caused large-scale epidemic outbreaks in South America (78). In combination with other epidemiological information, these data formed the basis of a decision to implement tetravalent A, C, W, Y conjugate vaccines into the UK immunization program for teenagers (79). Certainly, this was the first use of genomics to change national vaccination policy for meningococci and perhaps for any organism.

GENOME SEQUENCING VACCINE DESIGN AND ASSESSMENT

The availability of whole genome sequences (WGS) of bacterial pathogens also provided novel opportunities in the development of vaccines. In the case of the meningococcus, the first meningococcal WGSs, from bacterial isolates MC58 (80) and

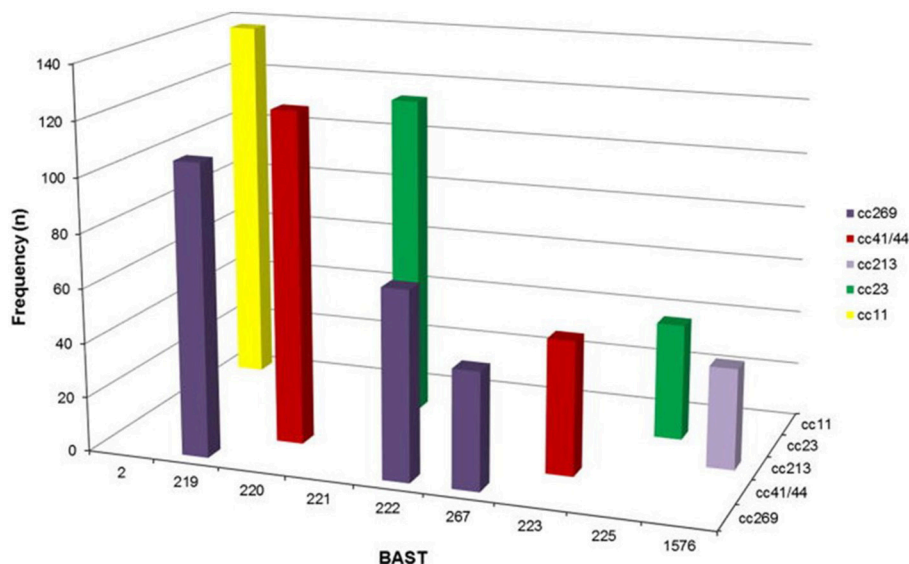


FIGURE 6 | Frequency distribution of Bexsero Antigen Sequence Typing (BAST) by clonal complex for the nine most frequently occurring BASTs ($n = 775/2,016$, 39.4%) in invasive meningococcal disease isolates in the epidemiological years 2010/2011 to 2013/2014. BASTs shown on the x axis are structured by clonal complex (z axis) for a proportion of frequently occurring dominant clones circulating in Great Britain and Ireland from 2010/2014, for example BAST-2 is only found in isolates of clonal complex 269 has three predominant BASTs, 219, 222, and 267 not found in other clonal complexes. Reproduced without modification from Brehony et al. (91) under CC BY.

Z2491 (81) both published 2000 played a role in the development of two “serogroup B substitute” vaccines: Bexsero[®] (4CMenB, developed in Siena, Italy) (82); and Trumenba[®] (LP2086, developed in Pearl River, USA) (83). In the case of the Bexsero[®] vaccine a “reverse vaccinology” (84) approach was adopted using the MC58 genome sequences as its starting point. In contrast to more conventional approaches, which took the interaction of a bacterial isolate as a starting point, reverse vaccinology started by predicting potential vaccine antigens (i.e., surface-exposed proteins) from a genome sequence, and then assessing these in animals. This identified three potential targets, fHbp, NadA, and NHBA that were eventually included in the final Bexsero[®] formulation, along with the MenNZB OMV vaccine, made against the New Zealand outbreak strain (38). Interestingly, a more conventional vaccine development approach, but which did use sequence information from the Z2498 meningococcal isolate to find the gene sequence from protein sequences, identified the fHbp protein (also known as LP2086) as an important vaccine candidate (83).

Given the known diversity of meningococcal protein antigens, assessment of the levels of diversity of these new vaccine components, especially the leading candidate fHbp (LP2086) formed a major part of the preclinical studies. Based on the analysis of deduced peptide sequences, the Pearl River group (studying “LP086”) proposed two subfamilies of the protein, A and B, whereas the Siena group (studying “GNA1870” later called fHbp), using a similar analysis of a different meningococcal collection, proposed three subvariants (1, 2, and 3, with subvariants 2 and 3 more closely related to each other) (85). As further sequencing was performed of this antigen, it became necessary to cross-reference and unify the

sequence nomenclatures between the two typing schemes and a single nomenclature was proposed (14) (Figure 5), along with a web accessible database enabling the cross-referencing of the various nomenclature schemes (<https://pubmlst.org/neisseria/fHbp/>). At the time of writing (September 2018), a total of 1,157 peptide sequence variants of this protein were described on the database.

As with the capsular polysaccharide-conjugate vaccines, it was impractical to conduct phase III efficacy studies on these novel “serogroup B substitute” vaccines. Further, the correlates of protection were less well-established than those for the capsular polysaccharide vaccines, where bactericidal assays using blood samples from vaccinees were employed (3). This was a particular problem in assessing the breadth of coverage of these vaccines, given the diversity of meningococcal protein antigens. This prompted the development of indirect assessments of coverage: the Meningococcal Antigen Typing System (MATS) assay for Bexsero[®], which incorporates sequence data from the *porA* gene (86); and the Meningococcal Antigen Surface Expression (MEASURE) Assay for Trumenba[®] (87). Both vaccines were licensed and used based on phase II clinical studies and studies using these assays (88, 89).

Pre- and post-introduction assessment of the coverage of these vaccines is an important component of assessing their likely and continued efficacy, especially for an organism as diverse and dynamic as the meningococcus. An efficient way to achieve this is by the extraction of antigen gene sequences from genome data collected as part of routine surveillance. The PubMLST website, which hosts data for the MRF-MGL and a number of other important global isolate collections, indexes all known genes and these can be flexibly grouped into “schemes,” which are

groups of genetic loci that are analyzed together for typing or functional purposes (13, 90). It was straightforward to combine the typing schemes for the various antigens in Bexsero[®] into a Bexsero[®] Vaccine Antigen Typing Scheme (BAST) (91), which enabled the assessment of the changing prevalence of the vaccine antigen variants in the UK and Ireland (91, 92) (Figure 6). Approaches such as this have great potential for supporting vaccine development implementation and monitoring into the future.

FUTURE PROSPECTS

As demonstrated by the examples outlined above, nucleotide sequence data, ranging from single gene fragments from individual bacterial specimens to whole completed genomes that are representative of populations, have many applications in vaccinology. These data are particularly useful in the rapid and cost-effective characterization of bacterial isolates. Combining such data with population and evolutionary analyses generates many informative inferences; however, whilst this complements data on the interaction of bacterial components with the host immune system, sequence analyses cannot wholly replace immunological studies. In an era where nucleotide sequences are low-cost commodities, the important advances of the future will depend on the interpretation and open-access dissemination of these data. In addition to novel statistical genetic techniques and integration with phenotypic data, the implementation of visualization tools is likely to be important in the further exploitation of this rich source of biological information.

REFERENCES

- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* (1977) 74:5463–7. doi: 10.1073/pnas.74.12.5463
- Maiden MCJ. The impact of protein-conjugate polysaccharide vaccines: an endgame for meningitis? *Philos T R Soc B* (2013) 368:20120147. doi: 10.1098/rstb.2012.0147
- Feavers IM, Maiden MC. Recent progress in the prevention of serogroup B meningococcal disease. *Clin Vaccine Immunol.* (2017) 24:e00566–16. doi: 10.1128/CVI.00566-16
- Finne J, Leinonen M, Makela PH. Antigenic similarities between brain components and bacteria causing meningitis. Implications for vaccine development and pathogenesis. *Lancet* (1983) 2:355–7. doi: 10.1016/S0140-6736(83)90340-9
- Poolman JT, Hopman CTP, Zanen HC. Immunogenicity of meningococcal antigens as detected in patient sera. *Infect Immun.* (1983) 40:398–406.
- Poolman JT, Beuvery EC, Hopman CTP, Witvliet MH, Timmermans HAM, Teerlink T, et al. Class 1/3 outer membrane protein vaccine against Group B, Type 15, Subtype 16 Meningococci. *Dev Biol Stand.* (1986) 63:147–52.
- Barlow AK, Heckels JE, Clarke IN. The class 1 outer membrane protein of *Neisseria meningitidis*: gene sequence and structural and immunological similarities to gonococcal porins. *Mol Microbiol.* (1989) 3:131–9. doi: 10.1111/j.1365-2958.1989.tb01802.x
- Maiden MCJ, Suker J, McKenna AJ, Bygraves JA, Feavers IM. Comparison of the class 1 outer membrane proteins of eight serological reference strains of *Neisseria meningitidis*. *Mol Microbiol.* (1991) 5:727–36. doi: 10.1111/j.1365-2958.1991.tb00743.x
- McGuinness B, Barlow AK, Clarke IN, Farley JE, Anilionis A, Poolman JT, et al. Deduced amino acid sequences of class 1 protein (PorA) from three strains of *Neisseria meningitidis*. Synthetic peptides define the epitopes responsible for serosubtype specificity. *J Exp Med.* (1990) 171:1871–82. doi: 10.1084/jem.171.6.1871
- van der Ley P, Heckels JE, Virji M, Hoogerhout P, Poolman JT. Topology of outer membrane porins in pathogenic *Neisseria* species. *Infect Immun.* (1991) 59:2963–71.
- Frasch CE, Zollinger WD, Poolman JT. Serotype antigens of *Neisseria meningitidis* and a proposed scheme for designation of serotypes. *Rev Infect Dis.* (1985) 7:504–10. doi: 10.1093/clinids/7.4.504
- Russell JE, Jolley KA, Feavers IM, Maiden MC, Suker J. PorA variable regions of *Neisseria meningitidis*. *Emerg Infect Dis.* (2004) 10:674–8. doi: 10.3201/eid1004.030247
- Jolley KA, Maiden MC. Automated extraction of typing information for bacterial pathogens from whole genome sequence data: *Neisseria meningitidis* as an exemplar. *Eurosurveillance* (2013) 18:20379. doi: 10.2807/ese.18.04.20379-en
- Brehony C, Wilson DJ, Maiden MC. Variation of the factor H-binding protein of *Neisseria meningitidis*. *Microbiology* (2009) 155:4155–69. doi: 10.1099/mic.0.027995-0
- Thompson EAL, Feavers IM, Maiden MCJ. Antigenic diversity of meningococcal enterobactin receptor FetA, a vaccine component. *Microbiology* (2003) 149(Pt 7):1849–58. doi: 10.1099/mic.0.26131-0
- Bambini S, De Chiara M, Muzzi A, Mora M, Lucidarme J, Brehony C, et al. *Neisseria* adhesin A variation and revised nomenclature scheme. *Clin Vaccine Immunol.* (2014) 21:966–971. doi: 10.1128/CVI.00825-13

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

During the earlier parts of this work (1988–1997) MM was employed by the UK National Health Service at the National Institute for Biological Standards and Control (NIBSC), he was funded for a sabbatical in the laboratory of Mark Achtman by the Alexander von Humboldt Stiftung (1996–7) from 1997 to 2015 he was Wellcome Trust Senior Fellow and from 2004-present Professor of Molecular Epidemiology and a Fellow of Hertford College at the University of Oxford. He is grateful to the European Union (especially grants QLK2-CT-2001-01436 and FP7-278864-2), the Wellcome Trust (especially grants: 047072/Z/96/B, 047072/Z/96/C, 062057/Z/00/Z, 081494/Z/06/Z, 087622/Z/08/A, 086546/Z/08/Z, 091634/Z/10/Z, and 104992/Z/14/Z), the Meningitis Research Foundation, the UK Department of Health (especially contract PR-ST-0915-10015) and the Oxford Martin School for funding.

ACKNOWLEDGMENTS

I would like to express my thanks to the members of past and present MaidenLab and the many collaborators who have contributed to my work over the years. Special thanks go to Ian Feavers of NIBSC, whom was the crucial collaborator in establishing this program of work.

17. Harrison OB, Evans NJ, Blair JM, Grimes HS, Tinsley CR, Nassif X, et al. Epidemiological evidence for the role of the hemoglobin receptor, HmbR, in meningococcal virulence. *J Infect Dis.* (2009) 200:94–8. doi: 10.1086/599377
18. van den Elsen JM, Herron JN, Hoogerhout P, Poolman JT, Boel E, Logtenberg T, et al. Bactericidal antibody recognition of a PorA epitope of *Neisseria meningitidis*: crystal structure of a Fab fragment in complex with a fluorescein-conjugated peptide. *Proteins* (1997) 29:113–25.
19. Tzitzilonis C, Prince SM, Collins RF, Achtman M, Feavers IM, Maiden MC, et al. Structural variation and immune recognition of the P1.2 subtype meningococcal antigen. *Proteins* (2006) 62:947–55. doi: 10.1002/prot.20800
20. Saleem M, Prince SM, Rigby SE, Imran M, Patel H, Chan H, et al. Use of a molecular decoy to segregate transport from antigenicity in the FrpB iron transporter from *Neisseria meningitidis*. *PLoS ONE* (2013) 8:e56746. doi: 10.1371/journal.pone.0056746
21. Moxon ER, Rainey PB, Nowak MA, Lenski RE. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol.* (1994) 4:24–33. doi: 10.1016/S0960-9822(00)00005-1
22. Feavers IM, Heath AB, Bygraves JA, Maiden MCJ. Role of horizontal genetic exchange in the antigenic variation of the class 1 outer membrane protein of *Neisseria meningitidis*. *Mol Microbiol.* (1992) 6:489–95. doi: 10.1111/j.1365-2958.1992.tb01493.x
23. Spratt BG. Hybrid penicillin-binding proteins in penicillin-resistant strains of *Neisseria gonorrhoeae*. *Nature* (1988) 332:173–6. doi: 10.1038/332173a0
24. Dowson CG, Hutchison A, Brannigan JA, George RC, Hansman D, Liñares J, et al. Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Proc Natl Acad Sci USA* (1989) 86:8842–6. doi: 10.1073/pnas.86.22.8842
25. Maynard Smith J, Dowson CG, Spratt BG. Localized sex in bacteria. *Nature* (1991) 349:29–31. doi: 10.1038/349029a0
26. Selander RK, Caugant DA, Ochman H, Musser JM, Gilmour MN, Whittam TS. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol.* (1986) 51:837–84.
27. Wang JF, Caugant DA, Morelli G, Koumaré B, Achtman M. Antigenic and epidemiological properties of the ET-37 complex of *Neisseria meningitidis*. *J Infect Dis.* (1993) 167:1320–9. doi: 10.1093/infdis/167.6.1320
28. Olyhoek T, Crowe BA, Achtman M. Clonal population structure of *Neisseria meningitidis* serogroup A isolated from epidemics and pandemics between 1915 and 1983. *Rev Infect Dis.* (1987) 9:665–82. doi: 10.1093/clinids/9.4.665
29. Caugant DA, Mocca LF, Frasch CE, Frøholm LO, Zollinger WD, Selander RK. Genetic structure of *Neisseria meningitidis* populations in relation to serogroup, serotype, and outer membrane protein pattern. *J Bacteriol.* (1987) 169:2781–92. doi: 10.1128/jb.169.6.2781-2792.1987
30. Caugant DA, Frøholm LO, Bovre K, Holten E, Frasch CE, Mocca LF, et al. Intercontinental spread of a genetically distinctive complex of clones of *Neisseria meningitidis* causing epidemic disease. *Proc Natl Acad Sci USA* (1986) 83:4927–31. doi: 10.1073/pnas.83.13.4927
31. Caugant DA, Zollinger WD, Mocca LF, Frasch CE, Whittam TS, Frøholm LO, et al. Genetic relationships and clonal population structure of serotype 2 strains of *Neisseria meningitidis*. *Infect Immun.* (1987) 55:1503–13.
32. Wang JF, Caugant DA, Li X, Hu X, Poolman JT, Crowe BA, et al. Clonal and antigenic analysis of serogroup A *Neisseria meningitidis* with particular reference to epidemiological features of epidemic meningitis in China. *Infect Immun.* (1992) 60:5267–5282.
33. Caugant DA, Kristiansen BE, Frøholm LO, Bovre K, Selander RK. Clonal diversity of *Neisseria meningitidis* from a population of asymptomatic carriers. *Infect Immun.* (1988) 56:2060–8.
34. Caugant DA, Høiby EA, Rosenqvist E, Frøholm LO, Selander RK. Transmission of *Neisseria meningitidis* among asymptomatic military recruits and antibody analysis. *Epidemiol Infect.* (1992) 109:241–53. doi: 10.1017/S0950268800050196
35. Maiden MCJ, Caugant DA. The population biology of *Neisseria meningitidis*: implications for meningococcal disease, epidemiology and control. In: Frosch M, Maiden MCJ, editors. *Handbook of Meningococcal Disease*. Weinheim: Wiley-VCH Verlag GmbH & Co KGaA (2006). p. 17–35. doi: 10.1002/3527608508.ch2
36. Bjune G, Høiby EA, Grønnesby JK, Arnesen O, Fredriksen JH, Halstensen A, et al. Effect of outer membrane vesicle vaccine against group B meningococcal disease in Norway. *Lancet* (1991) 338:1093–6. doi: 10.1016/0140-6736(91)91961-S
37. Sierra GVG, Campa HC, Varcacel NM, Garcia IL, Izquierdo PL, Sotolongo PF, et al. Vaccine against group B *Neisseria meningitidis*: protection trial and mass vaccination results in Cuba. *NIPH Ann.* (1991) 14:195–207.
38. Oster P, Lennon D, O'Hallahan J, Mulholland K, Reid S, Martin D. McNZB: a safe and highly immunogenic tailor-made vaccine against the New Zealand *Neisseria meningitidis* serogroup B disease epidemic strain. *Vaccine* (2005) 23:2191–6. doi: 10.1016/j.vaccine.2005.01.063
39. Halter R, Pohlner J, Meyer TF. Mosaic-like organization of IgA protease genes in *Neisseria gonorrhoeae* generated by horizontal genetic exchange *in vivo*. *EMBO J.* (1989) 8:2737–44. doi: 10.1002/j.1460-2075.1989.tb08415.x
40. Spratt BG, Zhang Q-Y, Jones DM, Hutchison A, Brannigan JA, Dowson CG. Recruitment of a penicillin-binding protein gene from *Neisseria flavescens* during the emergence of penicillin resistance in *Neisseria meningitidis*. *Proc Natl Acad Sci USA* (1989) 86:8988–92. doi: 10.1073/pnas.86.2.8988
41. Feil E, Zhou J, Maynard Smith J, Spratt BG. A comparison of the nucleotide sequences of the *adk* and *recA* genes of pathogenic and commensal *Neisseria* species: evidence for extensive interspecies recombination within *adk*. *J Mol Evol.* (1996) 43:631–40. doi: 10.1007/BF02202111
42. Maiden MC. Multilocus sequence typing of bacteria. *Annu Rev Microbiol.* (2006) 60:561–88. doi: 10.1146/annurev.micro.59.030804.121325
43. Harrison OB, Claus H, Jiang Y, Bennett JS, Bratcher HB, Jolley KA, et al. Description and nomenclature of *Neisseria meningitidis* capsule locus. *Emerg Infect Dis.* (2013) 19:566–73. doi: 10.3201/eid1904.111799
44. Jolley KA, Brehony C, Maiden MC. Molecular typing of meningococci: recommendations for target choice and nomenclature. *FEMS Microbiol Rev.* (2007) 31:89–96. doi: 10.1111/j.1574-6976.2006.00057.x
45. Miller E, Salisbury D, Ramsay M. Planning, registration, and implementation of an immunisation campaign against meningococcal serogroup C disease in the UK: a success story. *Vaccine* (2001) 20(Suppl. 1):S58–67. doi: 10.1016/S0264-410X(01)00299-7
46. Pollard AJ, Perrett KP, Beverley PC. Maintaining protection against invasive bacteria with protein-polysaccharide conjugate vaccines. *Nat Rev Immunol.* (2009) 9:212–20. doi: 10.1038/nri2494
47. Maiden MCJ, Spratt BG. Meningococcal conjugate vaccines: new opportunities and new challenges. *Lancet* (1999) 354:615–6. doi: 10.1016/S0140-6736(99)00252-4
48. Maiden MC, Ibarz-Pavon AB, Urwin R, Gray SJ, Andrews NJ, Clarke SC, et al. Impact of meningococcal serogroup c conjugate vaccines on carriage and herd immunity. *J Infect Dis.* (2008) 197:737–43. doi: 10.1086/527401
49. Maiden MC, Stuart JM, UK_Meningococcal_Carriage_Group. Carriage of serogroup C meningococci 1 year after meningococcal C conjugate polysaccharide vaccination. *Lancet* (2002) 359:1829–31. doi: 10.1016/S0140-6736(02)08679-8
50. Ibarz-Pavon AB, MacLennan J, Andrews NJ, Gray SJ, Urwin R, Clarke SC, et al. Changes in serogroup and genotype prevalence among carried meningococci in the United Kingdom during vaccine implementation. *J Infect Dis.* (2011) 204:1046–53. doi: 10.1093/infdis/jir466
51. Trotter CL, Maiden MC. Meningococcal vaccines and herd immunity: lessons learned from serogroup C conjugate vaccination programs. *Exp Rev Vaccines* (2009) 8:551–61. doi: 10.1586/erv.09.48
52. Trotter CL, Andrews NJ, Kaczmarski EB, Miller E, Ramsay ME. Effectiveness of meningococcal serogroup C conjugate vaccine 4 years after introduction. *Lancet* (2004) 364:365–7. doi: 10.1016/S0140-6736(04)16725-1
53. Lapeyssonnie L. La méningite cérébrospinale en Afrique. *Bull World Health Organ* (1963) 28:53–114.
54. Greenwood B. Manson lecture. Meningococcal meningitis in Africa. *Trans R Soc Trop Med Hyg.* (1999) 93:341–53. doi: 10.1016/S0035-9203(99)90106-2

55. Tiffay K, Jodar L, Kieny MP, Socquet M, LaForce FM. The evolution of the meningitis vaccine project. *Clin Infect Dis.* (2015) 61:S396–403. doi: 10.1093/cid/civ594
56. Sow SO, Okoko BJ, Diallo A, Viviani S, Borrow R, Carlone G, et al. Immunogenicity and safety of a meningococcal conjugate vaccine in Africans. *N Engl J Med.* (2011) 364:2293–304. doi: 10.1056/NEJMoa1003812
57. Ali O, Aseffa A, Bedru A, Lemma T, Moti T, Worku A, et al. Meningococcal carriage in the African meningitis belt. *Trop Med Int Health* (2013) 18:968–78. doi: 10.1111/tmi.12125
58. Trotter CL, Greenwood BM. Meningococcal carriage in the African meningitis belt. *Lancet Infect Dis.* (2007) 7:797–803. doi: 10.1016/S1473-3099(07)70288-8
59. MacLennan J, Kafatos G, Neal K, Andrews N, Cameron JC, Roberts R, et al. Social behavior and meningococcal carriage in British teenagers. *Emerg Infect Dis.* (2006) 12:950–7. doi: 10.3201/eid1206.051297
60. Ali O, Aseffa A, Bedru A, Lema T, Moti T, Tekletsion Y, et al. The diversity of meningococcal carriage across the african meningitis belt and the impact of vaccination with a group a meningococcal conjugate vaccine. *J Infect Dis.* (2015) 212:1298–307. doi: 10.1093/infdis/jiv211
61. Diallo K, Trotter C, Timbine Y, Tamboura B, Sow SO, Issaka B, et al. Pharyngeal carriage of *Neisseria* species in the African meningitis belt. *J Infect.* (2016) 72:667–77. doi: 10.1016/j.jinf.2016.03.010
62. Harmsen D, Singer C, Rothganger J, Tonjum T, de Hoog GS, Shah H, et al. Diagnostics of *neisseriaceae* and *moraxellaceae* by ribosomal DNA sequencing: ribosomal differentiation of medical microorganisms. *J Clin Microbiol.* (2001) 39:936–42. doi: 10.1128/JCM.39.3.936-942.2001
63. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony CM, Colles FM, et al. Ribosomal multi-locus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* (2012) 158:1005–15. doi: 10.1099/mic.0.055459-0
64. Bennett JS, Watkins ER, Jolley KA, Harrison OB, Maiden MC. Identifying *Neisseria* species using the 50S ribosomal protein L6 (*rplF*) gene. *J Clin Microbiol.* (2014) 52:1375–81. doi: 10.1128/JCM.03529-13
65. Christensen H, May M, Bowen L, Hickman M, Trotter CL. Meningococcal carriage by age: a systematic review and meta-analysis. *Lancet Infect Dis.* (2010) 10:853–61. doi: 10.1016/S1473-3099(10)70251-6
66. Ali O, Aseffa A, Omer AB, Lema T, Demissie TM, Tekletsion Y, et al. Household transmission of *Neisseria meningitidis* in the African meningitis belt: a longitudinal cohort study. *Lancet Glob Health* (2016) 4:E989–95. doi: 10.1016/S2214-109X(16)30244-3
67. Kristiansen PA, Diomande F, Ba AK, Sanou I, Ouedraogo AS, Ouedraogo R, et al. Impact of the serogroup a meningococcal conjugate vaccine, MenAfriVac, on carriage and herd immunity. *Clin Infect Dis.* (2013) 56:354–63. doi: 10.1093/cid/cis892
68. Daugla DM, Gami JP, Gamougam K, Naibei N, Mbainadji L, Narbe M, et al. Effect of a serogroup A meningococcal conjugate vaccine (PsA-TT) on serogroup A meningococcal meningitis and carriage in Chad: a community study [corrected]. *Lancet* (2014) 383:40–7. doi: 10.1016/S0140-6736(13)61612-8
69. Diallo K, Gamougam K, Daugla DM, Harrison OB, Bray JE, Caugant DA, et al. Hierarchical genomic analysis of carried and invasive serogroup A *Neisseria meningitidis* during the 2011 epidemic in Chad. *BMC Genom.* (2017) 18:398. doi: 10.1186/s12864-017-3789-0
70. Black S, Pizza M, Nisum M, Rappuoli R. Toward a meningitis-free world. *Sci Transl Med.* (2012) 4:123ps5. doi: 10.1126/scitranslmed.3003859
71. Chen WH, Neuzil KM, Boyce CR, Pasetti MF, Reymann MK, Martellet L, et al. Safety and immunogenicity of a pentavalent meningococcal conjugate vaccine containing serogroups A, C, Y, W, and X in healthy adults: a phase 1, single-centre, double-blind, randomised, controlled study. *Lancet Infect Dis.* (2018) 18:1088–96. doi: 10.1016/S1473-3099(18)30400-6
72. Bratcher HB, Corton C, Jolley KA, Parkhill J, Maiden MC. A gene-by-gene population genomics platform: *de novo* assembly, annotation and genealogical analysis of 108 representative *Neisseria meningitidis* genomes. *BMC Genome.* (2014) 15:1138. doi: 10.1186/1471-2164-15-1138
73. Hill DMC, Lucidarme J, Gray SJ, Newbold LS, Ure R, Brehony C, et al. Genomic epidemiology of age-associated meningococcal lineages in national surveillance: an observational cohort study. *Lancet Infect Dis.* (2015) 15:1420–8. doi: 10.1016/S1473-3099(15)00267-4
74. Ladhani SN, Beebejaun K, Lucidarme J, Campbell H, Gray S, Kaczmarek E, et al. Increase in endemic *Neisseria meningitidis* capsular group W sequence type 11 complex associated with severe invasive disease in england and wales. *Clin Infect Dis.* (2015) 60:578–85. doi: 10.1093/cid/ciu881
75. Taha MK, Achtman M, Alonso JM, Greenwood B, Ramsay M, Fox A, et al. Serogroup W135 meningococcal disease in Hajj pilgrims. *Lancet* (2000) 356:2159. doi: 10.1016/S0140-6736(00)03502-9
76. Mayer LW, Reeves MW, Al-Hamdan N, Sacchi CT, Taha MK, Ajello GW, et al. Outbreak of W135 meningococcal disease in 2000: not emergence of a new W135 strain but clonal expansion within the electrophoretic type-37 complex. *J Infect Dis.* (2002) 185:1596–605. doi: 10.1086/340414
77. Aguilera JF, Perrocheau A, Meffre C, Hahne S. Outbreak of serogroup W135 meningococcal disease after the Hajj pilgrimage, Europe, 2000. *Emerg Infect Dis.* (2002) 8:761–7. doi: 10.3201/eid0808.010422
78. Lucidarme J, Hill DM, Bratcher HB, Gray SJ, du Plessis M, Tsang RS, et al. Genomic resolution of an aggressive, widespread, diverse and expanding meningococcal serogroup B, C and W lineage. *J Infect.* (2015) 71:544–52. doi: 10.1016/j.jinf.2015.07.007
79. Ladhani SN, Ramsay M, Borrow R, Riordan A, Watson JM, Pollard AJ. Enter B and W: two new meningococcal vaccine programmes launched. *Arch Dis Child.* (2016) 101:91–5. doi: 10.1136/archdischild-2015-308928
80. Tettelin H, Saunders NJ, Heidelberg J, Jeffries AC, Nelson KE, Eisen JA, et al. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* (2000) 287:1809–15. doi: 10.1126/science.287.5459.1809
81. Parkhill J, Achtman M, James KD, Bentley SD, Churcher C, Klee SR, et al. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* (2000) 404:502–6. doi: 10.1038/35006655
82. Vernikos G, Medini D. Bexsero® chronicle. *Pathog Glob Health* (2014) 108:305–16. doi: 10.1179/2047773214Y.0000000162
83. Zlotnick GW, Jones TR, Liberator P, Hao L, Harris S, McNeil LK, et al. The discovery and development of a novel vaccine to protect against *Neisseria meningitidis* Serogroup B Disease. *Hum Vaccines Immunother.* (2015) 11:5–13. doi: 10.4161/hv.34293
84. Pizza M, Scarlato V, Masignani V, Giuliani MM, Arico B, Comanducci M, et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* (2000) 287:1816–20. doi: 10.1126/science.287.5459.1816
85. Masignani V, Comanducci M, Giuliani MM, Bambini S, Adu-Bobie J, Arico B, et al. Vaccination against *Neisseria meningitidis* using three variants of the lipoprotein GNA1870. *J Exp Med.* (2003) 197:789–99. doi: 10.1084/jem.20021911
86. Donnelly J, Medini D, Boccadifuoco G, Biolchi A, Ward J, Frasch C, et al. Qualitative and quantitative assessment of meningococcal antigens to evaluate the potential strain coverage of protein-based vaccines. *Proc Natl Acad Sci USA* (2010) 107:19490–5. doi: 10.1073/pnas.1013758107
87. Donald RG, Hawkins JC, Hao L, Liberator P, Jones TR, Harris SL, et al. Meningococcal serogroup B vaccines: estimating breadth of coverage. *Hum Vaccines Immunother.* (2017) 13:255–65. doi: 10.1080/21645515.2017.1264750
88. Vogel U, Taha MK, Vazquez JA, Findlow J, Claus H, Stefanelli P, et al. Predicted strain coverage of a meningococcal multicomponent vaccine (4CMenB) in Europe: a qualitative and quantitative assessment. *Lancet Infect Dis.* (2013) 13:416–25. doi: 10.1016/S1473-3099(13)70006-9
89. McNeil LK, Donald RGK, Gribenko A, French R, Lambert N, Harris SL, et al. Predicting the susceptibility of meningococcal serogroup b isolates to bactericidal antibodies elicited by Bivalent rLP2086, a novel prophylactic vaccine. *Mbio* (2018) 9:e00036–18. doi: 10.1128/mBio.00036-18

90. Jolley KA, Maiden MC. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinform.* (2010) 11:595. doi: 10.1186/1471-2105-11-595
91. Brehony C, Rodrigues CM, Borrow R, Smith A, Cunney R, Moxon ER, et al. Distribution of Bexsero® Antigen Sequence Types (BASTs) in invasive meningococcal disease isolates: Implications for immunisation. *Vaccine* (2016) 34:4690–7. doi: 10.1016/j.vaccine.2016.08.015
92. Rodrigues C, Brehony C, Borrow R, Smith A, Cunney R, Moxon ER, et al. Genomic surveillance and meningococcal group B vaccine coverage estimates after introduction of the vaccine into the national immunisation programme in the UK. *Lancet* (2017) 389:85. doi: 10.1016/S0140-6736(17)30481-6

Conflict of Interest Statement: As an employee of the University of Oxford, MM has, over the past 20 years, undertaken contract research and consultancy for, and has been paid expenses and honoraria by, companies involved in vaccine development including GSK, Chiron, Novartis, Wyeth, and Pfizer.

Copyright © 2019 Maiden. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Comparison of Open-Source Reverse Vaccinology Programs for Bacterial Vaccine Antigen Discovery

Mattia Dalsass^{1,2}, Alessandro Brozzi¹, Duccio Medini¹ and Rino Rappuoli^{1*}

¹ GlaxoSmithKline, Siena, Italy, ² Dipartimento di Scienze Cliniche e Biologiche, Università degli Studi di Torino, Turin, Italy

OPEN ACCESS

Edited by:

Alexandre Barbosa Reis,
Universidade Federal de Ouro Preto,
Brazil

Reviewed by:

Rory Cristiane Fortes De Brito,
Universidade Federal de Ouro Preto,
Brazil

Paola Massari,
Tufts University School of Medicine,
United States

*Correspondence:

Rino Rappuoli
rino.rappuoli@gsk.com

Specialty section:

This article was submitted to
Vaccines and Molecular Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 18 October 2018

Accepted: 15 January 2019

Published: 14 February 2019

Citation:

Dalsass M, Brozzi A, Medini D and
Rappuoli R (2019) Comparison of
Open-Source Reverse Vaccinology
Programs for Bacterial Vaccine
Antigen Discovery.
Front. Immunol. 10:113.
doi: 10.3389/fimmu.2019.00113

Reverse Vaccinology (RV) is a widely used approach to identify potential vaccine candidates (PVCs) by screening the proteome of a pathogen through computational analyses. Since its first application in Group B *meningococcus* (MenB) vaccine in early 1990's, several software programs have been developed implementing different flavors of the first RV protocol. However, there has been no comprehensive review to date on these different RV tools. We have compared six of these applications designed for bacterial vaccines (NERVE, Vaxign, VaxiJen, Jenner-predict, Bowman-Heinson, and VacSol) against a set of 11 pathogens for which a curated list of known bacterial protective antigens (BPAs) was available. We present results on: (1) the comparison of criteria and programs used for the selection of PVCs (2) computational runtime and (3) performances in terms of fraction of proteome identified as PVC, fraction and enrichment of BPA identified in the set of PVCs. This review demonstrates that none of the programs was able to recall 100% of the tested set of BPAs and that the output lists of proteins are in poor agreement suggesting in the process of prioritize vaccine candidates not to rely on a single RV tool response. Singularly the best balance in terms of fraction of a proteome predicted as good candidate and recall of BPAs has been observed by the machine-learning approach proposed by Bowman (1) and enhanced by Heinson (2). Even though more performing than the other approaches it shows the disadvantage of limited accessibility to non-experts users and strong dependence between results and *a-priori* training dataset composition. In conclusion we believe that to significantly enhance the performances of next RV methods further studies should focus on the enhancement of accuracy of the existing protein annotation tools and should leverage on the assets of machine-learning techniques applied to biological datasets expanded also through the incorporation and curation of bacterial proteins characterized by negative experimental results.

Keywords: reverse vaccinology (RV) programs, antigen, bacterial pathogens, potential vaccine candidates (PVCs), bacterial protective antigens (BPAs)

INTRODUCTION

Reverse Vaccinology (RV) is a genome-based approach developed for the first time in early 1990's by Rappuoli (3) to identify meningococcal protein vaccine candidates in Group B *meningococcus* (MenB). In its original conception, since antigens inducing humoral antibody response are primarily located in extracellular or outer membrane district, all the open reading frames extracted

from the genome sequence of MenB strain MC58 were screened to select proteins predicted to be surface exposed, secreted or lipoproteins.

RV approach has revolutionized vaccine development by adopting computerized screening of protein sequences from the pathogen as the first step of the process, to select a subset of promising antigens, aka potential vaccine candidates (PVCs) (Figure 1A).

RV offers two main advantages compared to traditional vaccine development approaches: (i) identification of candidate antigens without the need to grow the pathogen (ii) identification of any antigen independently by its purified quantity to be suitable for vaccine testing.

Proteins returned by RV methods are called throughout this review PVCs (Potential Vaccine Candidates). Other names given to the selected proteins are VCs (Vaccine Candidates), VTs (Vaccine Targets), PVCs (Protein Vaccine Candidates). PVCs identified by RV undergo *in-vitro* and *in-vivo* validation through experimental assays aimed at confirming their protective potential. Each pathogen has its specific experimental assays and it is hard to standardize a common set of experimental features; the most common experimental evidences are the protection outcomes in animal models against virulent bacterial challenge or results obtained from correlate to protection like the human bactericidal assay (4). In the context of this review we refer to any candidate protein that gave positive results in confirmatory preclinical experimental assays as BPAs (bacterial protective antigens). In the literature synonymous of BPAs are protective antigens (PAg), known antigens (KA), or known protective antigens (KPA). Lists of BPAs for different bacteria or viruses might be found in databases like Violin (Protegen) (5) or mining the literature. A comprehensive review of the main biological features characterizing BPAs deposited in Violin (Protegen) might be found in Ong et al. (6).

The First RV Protocol

The first RV protocol started with the prediction of all open reading frames from the genome of MenB (strain MC58), in total 2,158. These open reading frames were screened to search for homology to bacterial surface-associated proteins using FASTA (7) and PSI-BLAST program (8). Proteins with no hits found (hypothetical proteins) were analyzed by PSORT (9), SignalP (10), and TMPRED program (11) to search for putative lipoproteins, secreted proteins, outer membrane, or periplasmatic proteins.

Abbreviations: RV, reverse vaccinology; MenB, Neisseria meningitidis serogroup B, Meningococcus B; PVC, potential vaccine candidate; VC, vaccine candidate; VT, vaccine target; BPA, bacterial protective antigen; PAg, protective antigen; KA, known antigen; KPA, known protective antigen; NERVE, new enhanced reverse vaccinology environment; BLAST, basic local alignment search tool; PSI-BLAST, position-specific iterated basic local alignment search tool; SBA, serum bactericidal assay; DEG, database of essential genes; OMV, outer membrane vesicle; ACC, auto cross covariance; SVM, support vector machine; MHC, major histocompatibility complex; MHC, major histocompatibility complex class II; VFDB, virulence factor database; TP, true positive; TN, true negative; FP, false positive; FN, false negative; fHbp, meningococcal factor H binding protein; NadA, Neisseria adhesin A.

From the 2,158 proteins, 570 were selected as PVCs. Out of them 350 were successfully expressed in *Escherichia coli* and injected to immunize mice. Sera from immunized animals were screened in a serum bactericidal assay (SBA)—a correlate of protection against invasive meningococcal diseases—and proteins with negative results were discarded. Among the 28 proteins able to induce bactericidal activity, 5 candidates were selected for final formulation and, combined to outer membrane vesicles (OMVs), later approved with the commercial name of *Bexsero*® (12).

RV Programs Overview

In the following years the RV protocol was successfully applied to other bacterial pathogens. These pathogens include *Chlamydia pneumonia* (13), *Streptococcus pneumoniae* (14) in which open reading frames encoding putative surface proteins and with significant homology to virulence factors of other bacteria were selected and *Porphyromonas gingivalis* (15), in which PVCs were identified by searching for global homology to proteins of known surface exposure or virulence. In these cases, the selection criteria to identify PVCs were restricted to extracellular subcellular localization and to homology to virulence factors already known in other bacterial species. A review about these first applications might be found in Masignani et al. (16).

Only in 2006 the first standalone RV program, distributed with the name of NERVE (New Enhanced reverse Vaccinology Environment), was published (17). Since then several other pathogen-independent RV programs have been released.

Until now there has been no comprehensive review of the available open-source RV programs and a systematic comparison on a benchmark dataset was missing. In this review we compared 6 open-source standalone RV programs designed for bacterial pathogens: NERVE, VaxiJen (18), Vaxign (19), Bowman-Heinson (1, 2), Jenner-predict (20), and VacSol (21). We tested them on 11 different bacterial proteomes.

RV Programs Categories

RV packages can be categorized in two types, according to their algorithmic approach: decision-tree or “filtering” and machine learning or “classifying.” Both types take as input protein sequences and call them as PVCs or not-PVCs.

Decision-tree or filtering RV programs

They are flowchart-like programs: the pathogen's protein sequences are passed through a series of filters until a subset is identified as PVCs. The filters are done on protein features that can be directly measured, like the molecular weight, or predicted by a computational program, like the subcellular localization or the probability to be an adhesion protein. When the filter is applied on a numerical feature (e.g., number of predicted transmembrane domains) an *a-priori* cut-off is used. Decision tree RV programs differ each other by the number of filters adopted. Examples of decision tree RV tools are NERVE (17), Vaxign (19), Jenner-predict (20), and VacSol (21).

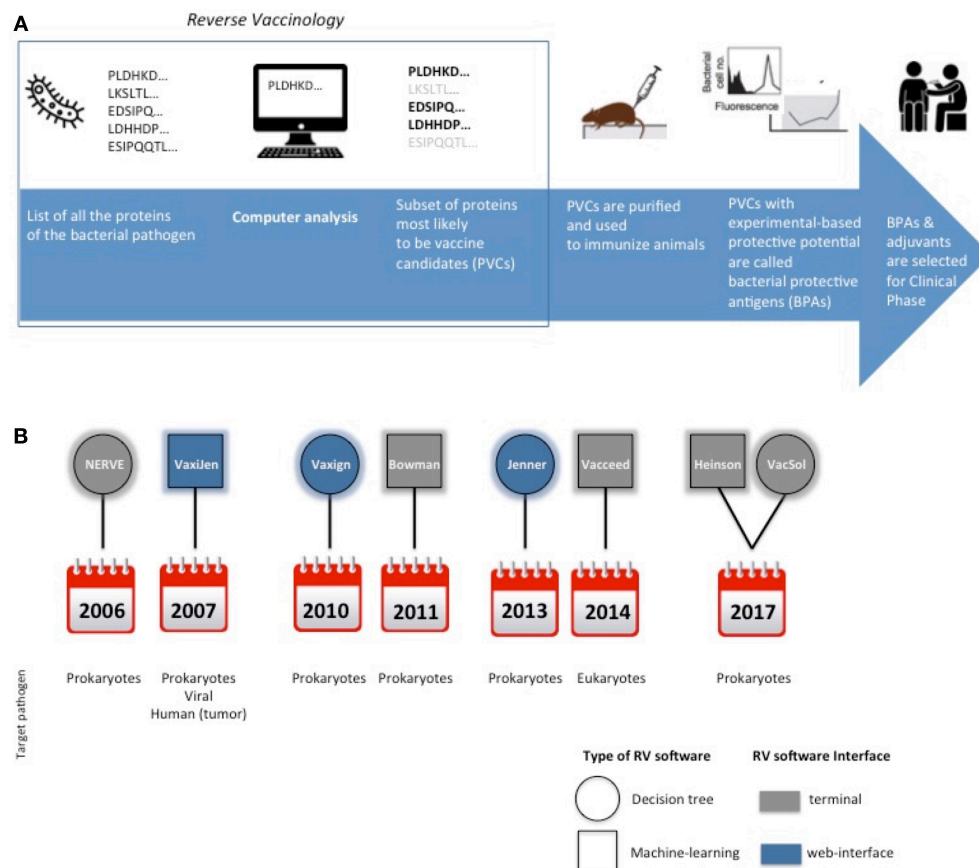


FIGURE 1 | Cartoon schematically representing the main steps for protein subunit vaccines development. In the square is highlighted the Reverse Vaccinology part (A). Timeline of the delivery of RV standalone programs and their main characteristics in terms of type of software, interface and target pathogen (B).

Machine-learning or classifying RV programs

These kinds of applications earlier aggregate the features measured or predicted on the pathogens' protein sequences into a matrix and then, given a known set of training examples of PVCs and not-PVCs, an algorithm builds a model that assigns new input proteins to one of the two classes usually in a probabilistic way. Machine-learning RV programs don't discard proteins, like decision tree RV tools do, but rank the entire set of input proteins for their likelihood of being a PVC. This results to be very useful when preclinical confirmatory assays must be planned since the experimenter might begin with the most promising candidates ranked in top positions.

Machine-learning RV tools are newer in the field and better intercept the increasing attention data analytics is paying to artificial intelligence methods. RV machine-learning tools differ each other from the type of classification algorithm they use, from the number of features they measure and from the size and assortment of proteins that constitute the training set.

Examples of machine-learning RV tools are VaxiJen (18), Vacceed (22) -designed for eukaryotes pathogens- and the method described in Bowman et al. (1) and revised by Heinson et al. (2).

Programs Interface

The interfaces to the RV programs fall into two categories, those that operate on the command line and those that have a graphical interface.

Command line input allows for high throughput analysis but has a high barrier to entry for non-technical users. Graphical interfaces, such as web-sites, provide point and click interfaces that non-technical users find easier to use initially, however, they are often limited to the analysis of a few samples at a time.

A synoptic summary of the types, year of release and interfaces of the six programs is provided in **Figure 1B**.

Software Description

In this section we describe one by one each of the six RV programs object of study of this review. We refer the reader to each specific publication for any further details. **Table 1** summarizes the criteria used by each of the six programs to identify PVCs and reports main advantages or disadvantages come upon their usage.

NERVE (17)

NERVE (New Enhanced Reverse Vaccinology Environment) has the primacy to be the first RV standalone software. It is a decision

TABLE 1 | Synoptic summary of the main characteristics of the six programs tested.

Tool	Category	PVC selection criterion	Advantage	Disadvantage	Number of studies that used the tool
NERVE	Decision-tree	No cytoplasmatic protein <2 transmembrane helices High adhesin probability No homology with human proteins	Input and output data are automatically structured in a database	Not updated	4
VaxiJen	Machine-learning	Output probability greater than a cut-off (0.5)	Very fast Graphical interface	Fixed training datasets (100 known bacterial antigens, 100 putative non-antigens)	20
Vaxign	Decision-tree	No cytoplasmatic protein <2 transmembrane helices High adhesin probability No homology with human and mouse proteins	Regularly maintained Easy to use and intuitive	Download of the results is limited to 500 proteins	18
Jenner-predict	Decision-tree	No cytoplasmatic protein <2 transmembrane helices Presence of Pfam domains involved in host-pathogen interaction and pathogenesis	Upload and download of large datasets	Temporarily unavailable	1
Bowman-Heinson	Machine-learning	Output probability greater than a cut-off (0.5)	Larger training set (200 known bacterial antigens, 200 putative non-antigens)	Annotation tools for eukaryotes used for bacterial proteins Pipeline not delivered	0
VacSol	Decision-tree	No cytoplasmatic protein <2 transmembrane helices No homology with human proteins Essential gene Virulence factor	User-friendly interface	Too restrictive	0

tree and command line tool. Once installed in a Unix-like operating system (NERVE is implemented in Perl programming language), the tool imports the sequences of the pathogen proteins and launches computational programs to predict five biological features:

- 1) subcellular localization (pSORT) (23)
- 2) adhesion probability (SPAAN) (24)
- 3) topology (HMMTOP) (25)
- 4) sequence similarity with human proteins (BLASTp) (8)
- 5) conservation in other strain of the same species (BLASTp).

NERVE parses the results of the five programs and stores the results in a MySQL database.

NERVE uses *a priori* cut-offs to select the PVCs. Based on tests done on 10 proteomes (*Bacillus anthracis*, *Pseudomonas aeruginosa*, *Yersinia pestis*, *Streptococcus agalactiae* V, III, Ia, *Neisseria meningitidis* B, *Porphyromonas gingivalis*, *Borrelia burgdorferi*, *Chlamydia trachomatis* D) the authors of NERVE suggest the following criterion to identify PVCs: any non-cytoplasmatic protein, with no more than 2 predicted transmembrane helices, with a predicted probability of being and adhesin >0.46 or 0.38 and without sequence similarity to human proteins.

NERVE shows the advantage to be very simple and intuitive; it also allows the user to change the filtering cut-offs according to his/her preferences for long or short lists of PVCs.

NERVE has not been updated since its first release: some Perl libraries became obsolete and to be used not negligible changes

must be done to the source code. Homology with human proteins is done comparing by BLAST algorithm each pathogen protein sequence against a dataset of potential MCH ligands derived from the database MHCPEP (26) that has not been updated since 1998.

VaxiJen (18)

Published soon after NERVE, is the first RV software adopting machine learning strategy. VaxiJen proposes an alignment-independent method for antigen prediction based on auto cross covariance (ACC) transformation of protein sequences into uniform equal-length vectors. Differently by other RV programs, VaxiJen might predict not only bacterial but also viral and tumor antigens. For bacterial antigens prediction VaxiJen applies ACC transformation to a set of 100 known bacterial antigens that the authors derived mining the literature; a protein was included in the set of known bacterial antigens if it (or part of it) was shown to induce a protective response in an appropriate animal model after immunization. Conversely a twin-set of 100 non-antigens was constructed to mirror the antigen set, randomly selecting proteins from the same set of species without similarity to the set of the 100 known antigens (BLAST expectation value of 3.0 was used). Two-class discriminant analysis by partial least squares was applied to the merged set (200 proteins) to derive a model of prediction that the user might apply on his own dataset of proteins uploading a file through a web-interface.

VaxiJen is a web-interface program. The results page reports antigen probability (as a fraction of unity) for each protein.

Criterion to call PVCs is any protein with an antigen probability above a threshold (defaults 0.5).

VaxiJen is the only tool currently allowing classification based solely on the physiochemical properties of protein sequences without any related biological or functional information.

While very easy to use and very fast a major limitation is though represented by the fact that, at least in its current release, it is not possible for the user to change the training dataset upon which the prediction model is derived. A review of VaxiJen applications during the last years might be found in Zaharieva et al. (27).

Vaxign (19)

Vaxign is decision-tree software that works via web-interface. Vaxign is available in two forms: *Vaxign Query* that provides precomputed results for users to explore, and *Dynamic Vaxign Analysis* that allows dynamic execution and result visualization.

In *Dynamic Vaxign Analysis*, likewise NERVE, it runs different external computational programs on input protein sequences to predict five biological properties:

- 1) Subcellular localization (pSORT) (23).
- 2) Number of transmembrane helices (HMMTOP) (25).
- 3) Adhesin probability (SPAAN) (24).
- 4) Similarity to host (human, mouse, pig) proteins (OrthoMCL) (28).
- 5) MHC I and MHC II epitopes binding (Vaxitope) (19).

The authors, analyzing 11 known protective antigens from four bacterial pathogens strains (*N. meningitidis*, *H. pylori*, *B. anthracis*, *M. tuberculosis*), suggest the following criterion to identify PVCs: any protein surface exposed, with no more than one transmembrane helix, with probability to be an adhesin >0.51 and no sequence similarity to any host protein (human and mouse).

Vaxign mostly resembles NERVE in terms of the protein features predicted, computational programs used and thresholds set to call PVCs, though there are differences:

- They both use pSORT to predict subcellular localization, but NERVE parses the original output and in case the probability to be “Cytoplasmatic” is null it reports “Non-cytoplasmatic.” On the contrary Vaxign keeps the original output of Psort.
- NERVE to call a PVC sets two different thresholds at the probability predicted by SPAAN of being an adhesin: 0.46 if the candidate is predicted “not cytoplasmatic” and 0.38 if predicted “extracellular.”
- Vaxign uses OrthoMCL (28) to calculate the homology to host proteins and the conservation on different strains of the pathogen, whereas NERVE implements a BLASTp query against a not updated dataset of MCHI binding epitopes derived from MHCPEP (26). NERVE filters PVCs for sequence similarity only against human proteins while Vaxign allows also against mouse and pig proteins.

Jenner-Predict (20)

It is decision-tree software published in 2013. Jenner-predict identifies PVCs by filtering upon:

- 1) Subcellular localization (pSORT, version 3.0) (29).
- 2) Number of transmembrane helices (HMMTOP, version 2.0) (25).
- 3) Presence of Pfam domains involved in host-pathogen interactions and pathogenesis (30).

Pfam domains include classes of adhesion, invasion, toxin, porins, colonization, virulence, flagellin, penicillin-binding, choline-binding, transferring-binding, fibronectin-binding, and solute-binding.

The criterion to identify PVCs for Jenner-predict is: any non-cytosolic protein with <3 trans-membrane helices and with at least one hit in the list of Pfam domains involved in host-pathogen interactions and pathogenesis. This final list of PVCs is then ranked according to the degree of conservation in different pathogenic and non-pathogenic strains, presence of known epitope sequences (both B and T epitopes) and degree of conservation with human proteins.

The novelty of Jenner-predict is to relax the criterion applied by NERVE and Vaxign on adhesin-likeness to call PVCs. Jenner-predict doesn't use SPAAN to predict the probability for a candidate to be an adhesin but uses Pfam domains.

Differently by Vaxign, Jenner-predict uses the sequence similarity to human proteins only as a score to rank PVC. Jenner-predict at the time of writing is unavailable for users through its web-interface. We contacted directly the authors to ask for a local evaluation of the software on our benchmark dataset.

Heinson-Bowman (1, 2)

We called this method with the names of the first authors who published a machine learning RV method initially in 2011 (1) and then enhanced the classifier publishing the results in 2017 (2). Bowman et al. (1) merged the existing tools NERVE and VaxiJen, adopting from NERVE the idea of use a set of protein annotation tools and from VaxiJen the use of a machine-learning classifier.

The method uses a Support Vector Machine (SVM) classifier using a training dataset constituted by 200 bacterial protective antigens (BPA) extracted from literature. Bacterial protective antigens mean that the proteins have evidences about their protective potential in an appropriate animal model after immunization. Other 200 non-BPA were randomly selected from the same proteomes without sharing sequence similarity to BPAs. This dataset was initially annotated with 525 features coming from 31 different annotation tools. After a feature selection step, the number of features has been reduced to 10. This short-list of 10 includes:

- 1) Average length of lipoprotein and other signal peptides (LipoP) (31).
- 2) Average length of signal peptide recognized by peptidase I (LipoP).
- 3) Count and length of O-beta-Glc-NAC attachment sites (YinOYang) (32).
- 4) Count of serine kinase specific eukaryotic phosphorylation sites (NetPhosK-S-Count) (33).
- 5) Average rank of human MHC alleles HLA-B matching the protein candidate (NetMhc) (34).

- 6) Presence of signal peptide for secretory pathway (TargetP) (35).
- 7) Count of I-Ag7 (MHC class II mice molecule) epitopes (GPS-MBA) (36).
- 8) Average score of MHC peptide binding sites (PickPocket) (37).
- 9) Count of scores of Furin-specific cleavage sites (ProP) (38).

The criterion to call PVC is any candidate protein with an antigen probability value greater than an a-priori threshold (0.5).

Even if the software is designed for bacterial PVCs, 8 out of 10 features are predicted by tools designed and tested for eukaryotic organisms, such as NetAcet (39) that predicts substrates of N-acetyl transferase A trained on yeast data with similar performances on mammalian substrates.

VacSol (21)

It is the last RV software appeared in the field of Reverse Vaccinology. It is decision-tree software that filters input protein candidates by:

- 1) Subcellular localization (PSORTb and CELLO2GO) (23, 40).
- 2) Similarity to host (human) proteins (BLASTp) (8).
- 3) Match to essential genes database (DEG) (41).
- 4) Match to virulence factors database (VFDB) (42).
- 5) Number of transmembrane helices (HMMTOP 2.0) (25).

The selection criterion for PVCs is: any non-host homologous, essential, virulent protein residing in the extracellular membrane with <2 transmembrane helices.

The final list of PVCs is then ranked accordingly to the prediction of MHC Class I and II binding regions and to the B-Cell epitope prediction.

BENCHMARK DATASETS

To compare the six RV tools, we selected a list of 11 bacterial species for which we could retrieve a list of BPAs combining information from literature (reviews) and publicly available in Protegen database (5).

The list of the 11 species -both Gram positive and Gram negative- includes bacteria that were already reported in the publications of the RV programs: eight species reported in NERVE publication or VacSol (*Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Staphylococcus aureus*, *Streptococcus pyogenes*, *Helicobacter pylori*, *Chlamydia pneumoniae*, *Campylobacter jejuni*, *Borrelia burgdorferi*), two reported in Jenner-predict publication or Vaxign (*Escherichia coli*, *Streptococcus pneumoniae*) and *Treponema pallidum*. For each species the list of BPAs and their relative references is reported in Table S1.

EVALUATION

Regarding Bowman-Heinson the original material has not been made available by the authors within the timelines needed to submit this manuscript. Being the pipeline of the program unavailable we decided to reproduce the analysis as far as possible in line with the description present in the articles (2).

TABLE 2 | Prototype of the golden-standard 2 x 2 table to measure the RV performances.

		Immunological assays readout	
		Positive (BPA)	Negative (no BPA)
RV method	PVC	True positive (TP)	False positive (FP)
prediction	Not-PVC	False negative (FN)	True negative (TN)

For each bacterial species the proteome was downloaded from Uniprot database (43) version 2018_05 and was given in input to each RV program that returned in output the list of PVCs. In not specified, default settings were used for each RV program.

Performances' Measures of RV Programs

The golden standard to measure how well a RV program performs would be in theory to purify all the pathogen's proteins, test experimentally each of them in the appropriate animal model through pathogen-specific laboratory assays and finally compare predictions and experimental results like in Table 2.

From results arranged like in Table 2 one could calculate both sensitivity or recall (TP/TP + FN), specificity (TN/TN + FP) and other performance metrics. Though in real-world scenario Table 2 is almost unfeasible because of time and cost constraints for entire bacterial proteomes that consists of thousands of proteins. In this review we decided to focus on BPAs only and accordingly to measure the performances of RV methods by:

- i) Fraction of proteome called PVCs (PVCs/proteome).
- ii) Fraction of BPA identified within the set of PVCs (sensitivity or recall).
- iii) Fold-enrichment expressed as ratio between number of BPAs observed in the set of PVCs and the number expected drawing from the proteome a random sample of the same size of the set of PVCs (statistical significance of the fold-enrichment assessed through an hypergeometric test).

RESULTS

Comparison of the PVC Selection Criteria and Computational Tools

VaxiJen classifies PVCs extracting information from the chemical-physical properties of the aminoacids composing bacterial proteins. Conversely the remaining five tools in order to define PVCs work on features predicted by external programs (for a list see Table 3).

From the comparison of the PVCs selecting criteria of these five RV programs we observed that they share two common features:

1. Extracellular subcellular localization
2. Probability of being a virulence factor

About the prediction of the extracellular localization the RV programs use mostly Psort while Bowman-Heinson implements TargetP.

TABLE 3 | Summary of the external computational programs used by the six programs to predict the protein features instrumental to filter or classify PVCs.

Protein feature	Prediction program	NERVE	Vaxign	Jenner-predict	VacSol	Bowman-Heinson
Subcellular localization	Psortb	X	X	X	X	
	TargetP					X
Transmembrane domains	HMMTOP	X	X	X	X	
Pathogenic domains or virulent factors	SPAAN	X	X			
	Pfam			X		
	VFDB				X	
	LipoP					X
Similarity to host proteins	BLASTp against MHCPEP db	X				
	BLASTp against RefSeq and Swiss Prot db				X	
	OrthoMCL		X			
B-T cell response	NetMhc					X
	Vaxitope		X			
	ABCPred				X	
	ProPred-I				X	
	ProPred				X	
	GPS-MBA					X
	PickPocket					X
						X
Post-translational modification	YinOYang (glycosylation)					X
	NetPhosK (phosphorylation)					X
	ProP (proprotein convertase cleavage)					X

The major virulence characteristic that is searched for is adhesion. SPAAN is the software of election to predict the probability of a protein being an adhesin and is used by NERVE and Vaxign. VacSol searches PVCs in the database of virulence factors VFDB that contains discrete proportion of adhesins. Among the Pfam domains used by Jenner-predict 96 domains are reported as related to adhesion. Also lipoproteins have been shown to play key roles in adhesion to host cells and translocation of virulence factors into host cells (44). Heinson uses LipoP software that produces predictions of lipoproteins.

Differently from what one might expect not all the RV programs use the sequence similarity to host proteins (either mouse or human) as a selecting criterion. Jenner-predict for instance uses the homology to human proteins only to rank the PVCs accordingly to what they call their “vaccine potential.” Machine-learning approach of Heinson doesn’t include in the list of the 525 initial potential discriminative features anything related to homology or similarity to host proteins.

Finally, HMMTOP is common to all the four decision-tree programs NERVE, Vaxign, Jenner-predict and VacSol. It is used to predict the number of transmembrane domains that is directly linked to the likelihood each protein has to be successfully purified.

Running Time

The performances in terms of time needed to predict PVCs are reported in **Table 4**. Time has been calculated using a set of 100 protein sequences with an average length of 360

TABLE 4 | Summary of run times on a benchmark dataset of 100 proteins (average length 360 a.a.).

Program	Running time
VaxiJen	5 s
Vaxign	5 min 40 s
NERVE	17 min 37 s
Bowman-Heinson	27 min 8 s
VacSol	49 min 40 s

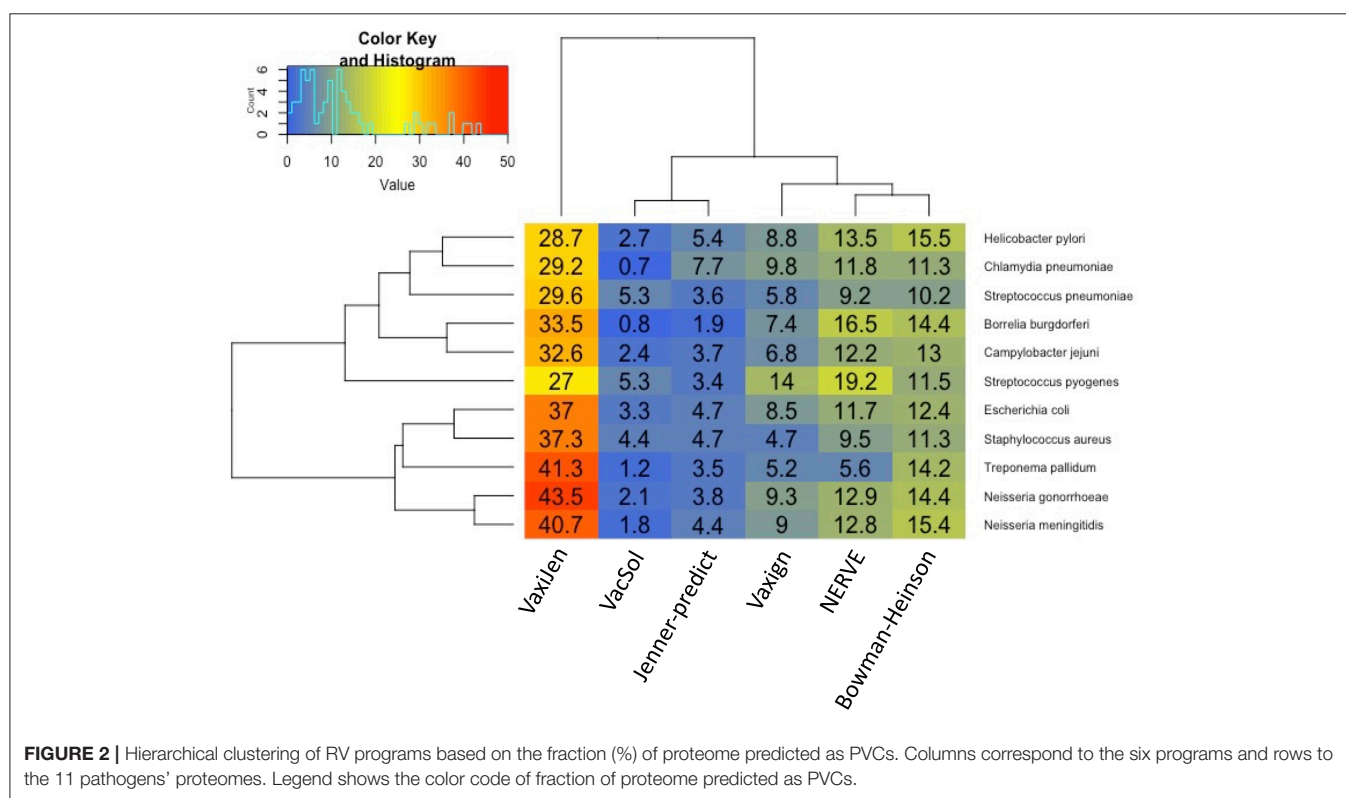
aminoacids. Tools like Vaxign and VaxiJen are very fast and are able to predict 100 proteins in a few seconds or minutes, instead other tools like NERVE, Bowman-Heinson and VacSol are slower and need between 15 and 60 min to analyze the same protein dataset on a MacBook Pro (2.6 GHz Intel Core i7, 16 Gb RAM).

This difference is due to the fact that tools used via browser like Vaxign and VaxiJen have been developed in a specific way integrating the software needs with the hardware. In the case of tools such as NERVE, VacSol and Bowman-Heinson, the analysis depends on the characteristics of the hardware used and the running time may vary depending on the capabilities of the system.

In addition, must be noticed that tools like NERVE, VacSol and Bowman-Heinson are not available as preconfigured virtual machine so time must be dedicated to install the software itself and all its dependencies. Vaxign and VaxiJen, available via

TABLE 5 | Fraction of PVCs predicted by each of the six programs (NERVE, VaxiJen, Vaxign, VacSol, Bowman-Heinson, and Jenner-predict) where pathogens are listed following the order of their proteome size.

Species	RV programs					
	NERVE	VaxiJen	Vaxign	VacSol	Bowman-Heinson	Jenner-predict
<i>Escherichia coli</i>	627 (11.7%)	1,979 (37%)	452 (8.5%)	174 (3.3%)	661 (12.4%)	250 (4.7%)
<i>Streptococcus pyogenes</i>	690 (19.2%)	972 (27%)	504 (14%)	190 (5.3%)	414 (11.5%)	121 (3.4%)
<i>Chlamydia pneumoniae</i>	398 (11.8%)	984 (29.2%)	330 (9.8%)	25 (0.7%)	380 (11.3%)	260 (7.7%)
<i>Staphylococcus aureus</i>	254 (9.5%)	992 (37.3%)	125 (4.7%)	118 (4.4%)	300 (11.3%)	126 (4.7%)
<i>Streptococcus pneumoniae</i>	194 (9.2%)	625 (29.6%)	122 (5.8%)	111 (5.3%)	216 (10.2%)	75 (3.6%)
<i>Neisseria gonorrhoeae</i>	272 (12.9%)	917 (43.5%)	197 (9.3%)	45 (2.1%)	304 (14.4%)	81 (3.8%)
<i>Neisseria meningitidis</i>	256 (12.8%)	815 (40.7%)	180 (9%)	37 (1.8%)	308 (15.4%)	88 (4.4%)
<i>Treponema pallidum</i>	92 (5.6%)	682 (41.3%)	85 (5.2%)	20 (1.2%)	234 (14.2%)	58 (3.5%)
<i>Campylobacter jejuni</i>	199 (12.2%)	530 (32.6%)	111 (6.8%)	39 (2.4%)	211 (13%)	60 (3.7%)
<i>Helicobacter pylori</i>	201 (13.5%)	429 (28.7%)	131 (8.8%)	40 (2.7%)	231 (15.5%)	81 (5.4%)
<i>Borrelia burgdorferi</i>	213 (16.5%)	432 (33.5%)	96 (7.4%)	10 (0.8%)	186 (14.4%)	25 (1.9%)
Average on 27,247 total proteins	3,396 (12.5%)	9,357 (34.3%)	2,602 (9.5%)	809 (3%)	3,445 (12.6%)	1,225 (4.5%)



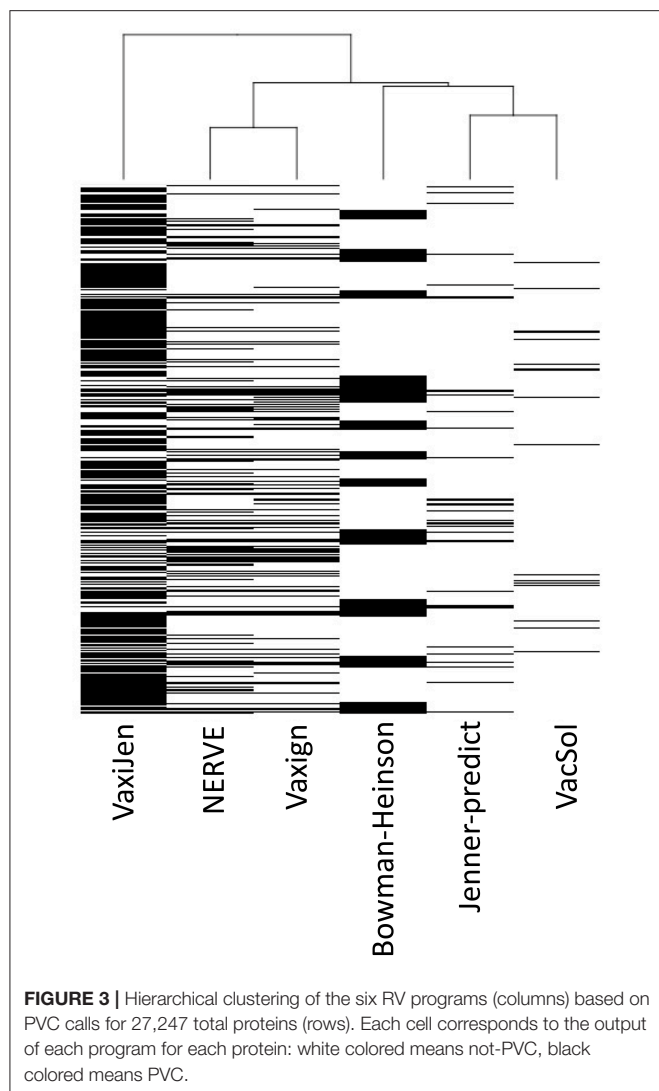
browser, are easier to use, only necessitating to copy and paste fasta sequences of the proteins.

Fraction of PVCs

The results are presented in Table 5 where pathogens are listed following the order of their proteome size (decreasing order).

Among the six programs VacSol resulted to be the most conservative predicting as PVCs on average only 3% of a

bacterial proteome (min 0.7% *Chlamydia pneumoniae*—max 5.3% *Streptococcus pyogenes*). On the opposite side VaxiJen is the most permissive with on average 34.4% (min 27% *Staphylococcus aureus*—max 43.5% *Neisseria gonorrhoeae*) of a bacterial proteome predicted as PVC. A graphical summary is provided in Figure 2. As shown in the figure based on proteome fraction predicted as PVC we could hierarchically cluster the six programs into three groups corresponding to high, medium and low fraction of predicted PVCs.



VaxiJen is the software that predicts the greatest fractions of PVCs (always more than 25% of a proteome) and stands separately from the other tools. VacSol and Jenner-predict constitute the second group with low fractions of PVCs (always <10% of a proteome). In the middle are NERVE, Vaxign and Bowman-Heinson with similar medium fractions of PVCs predicted.

Analyzing the output of the six RV programs for each single protein we observed heterogeneous agreement among the programs (Figure 3). To quantify the strength of each pair-wise agreement among the six programs we used the Choen's kappa (45). If two programs are in complete agreement, then kappa is equal to 1. If there is no agreement between two programs other than what would be expected by chance kappa is equal or even <0. The values of kappa for the pairwise comparisons between programs are given in Table 6.

The programs are scarcely in agreement with the only exception of NERVE and Vaxign that show a high kappa value

TABLE 6 | Choen's kappa values for the pair-wise agreement between programs.

	VacSol	VaxiJen	Vaxign	NERVE	Bowman-Heinson
VaxiJen	−0.012				
Vaxign	−0.016	0.129			
NERVE	−0.014	0.135	0.769		
Bowman-Heinson	0.009	0.126	0.278	0.276	
Jenner-predict	0.032	0.073	0.321	0.269	0.275

In bold the maximum value of each column.

(0.769). VacSol seems to be the software that returns a list of PVCs mostly not in common with others (kappa ranges between −0.012 and 0.032).

Fraction of BPAs Identified and Fold-Enrichment

For each software we measured the fraction of BPAs identified in the subset of PVCs, the recall and the fold-enrichment associated with *p*-value based on hypergeometric distribution as described in section Performances' Measures of RV Programs.

As reported in Table 7 the software with the highest fold enrichment is Jenner-predict that however has a recall of 44%. VaxiJen recalls the maximum absolute number of BPAs (76 BPAs in 9,357 PVCs) but has a low fold-enrichment (2.2). In comparison to VaxiJen, Bowman-Heinson with 3,445 PVCs recalls 75 BPAs showing therefore the best performance in terms of combined recall and fold-enrichment (5.9). Data for each single pathogen are provided in Table S2.

DISCUSSION

Reverse vaccinology represents a critical step toward the discovery and development of protein subunit vaccines.

From its conception in early 2000 to date several programs have been developed to do Reverse Vaccinology. We reviewed six of them, open-source, designed for bacterial pathogens.

We found two types of RV programs: those based on decision-tree or filtering and those based on machine-learning or classifying.

The first type—including NERVE, Jenner-predict, Vaxign, and VacSol—has the advantage of using a predefined set of core features to predict PVCs, without requiring training on a preexisting list of good and bad candidates. Core features include extracellular localization, probability to be an adhesion, lack of similarity to host proteins and limited number of transmembrane domains.

We observed that on average 10–15% of a bacterial proteome matches these criteria, resulting in a list of hundreds of proteins to be potentially tested in preclinical laboratories.

Conversely, methods based on machine-learning use training sets. VaxiJen uses as predictive features values calculated from the aminoacidic composition of the proteins and returns long lists of PVCs: on average one third of a bacterial

TABLE 7 | Summary of the performance on the RV programs in terms of recall of BPAs and fold-enrichment.

Software	PVCs	Observed BPAs	Recall (%)	Expected BPAs	Fold-enrichment	p-Value
NERVE	3,396	64	64	12	5.1	1.51E-33
VaxiJen	9,357	76	76	34	2.2	1.80E-17
Vaxign	2,602	58	58	10	6.1	1.90E-33
VacSol	809	4	4	3	1.3	3.46E-01
Bowman-Heinson	3,445	75	75	13	5.9	1.99E-46
Jenner-predict	1,225	44	44	4	9.8	1.09E-32

In bold the maximum value of each column. Numbers are referred to the total number of proteins (27,247) of the 11 pathogens. BPAs are 100 in total.

proteome is called PVC. It is likely that changing the training set—at the time of writing the review composed by 200 proteins—the output lists of PVCs might change as well. The other machine-learning approach (based on a Support Vector Machine) developed by Bowman and enhanced by Heinson uses features extracted from programs predictive of subcellular localization, B and T cell responses and post-translational modifications. Differently by VaxiJen the output list of PVCs is contained (12% of a proteome on average) and the method shows in our benchmark dataset a valuable enrichment in BPAs.

One advantage of the filtering RV programs is represented by user's full control of the step-wise process toward the selection of PVCs. PVCs are then easy to interpret and communicate. NERVE has not been updated since its release though Vaxign constitutes a valid alternative as it implements a very similar pipeline. The accordance between the two is indeed very good. VacSol represents also a valid RV filtering program but the number of resulting PVCs is so restricted that the likelihood to miss good candidates is not negligible.

Machine-learning methods are able to rank all the proteins of a pathogen based on their likelihood to be a PVCs. They can handle simultaneously much more features than filtering RV methods. However, these methods need an *a-priori* training dataset of good and bad antigens. This represents their main Achilles' heel because if it true that on literature one might find experimental evidences for good antigens, the same is not always valid for negative cases i.e., candidate proteins that didn't succeed in preclinical testing. The shortcut commonly used to artificially populate a set of bad antigens randomly selecting proteins not tested in laboratory but with scarce similarity to good antigens is questionable. Evidence of this are for instance the two antigens fHbp and NadA present in *Bexsero*[®] vaccine. Considering fHbp a good antigen, based on the almost null sequence similarity to NadA one would consider NadA as bad candidate. It would be beneficial to increase the performances of RV methods if manually curated set of candidate proteins with negative experimental outcomes would be publicly available. A limitation of machine-learning RV methods might be represented by the interpretation of the results since it is not straightforward to map backwardly PVCs to the features space.

CONCLUSIONS

We have extensively reviewed, for the first time, the state-of-the-art of Reverse Vaccinology bioinformatic tools used in bacterial antigen prioritization, visualized their diversity, and examined their performances.

We found that independently by the number of predicted PVCs, none of the six programs was able to recall more than 76 BPAs out of the benchmark list of 100 composed from eleven different bacterial species. The machine learning based method of Bowman-Heinson demonstrated the best ratio between BPA identified and number of PVCs predicted, recalling 75% of BPAs in a total of 3,445 PVCs. This is relevant in the field because while reducing the number of laboratory tests this method should simultaneously guarantee the identification of the vast majority of proteins with potential protective efficacy.

When we looked at the overall agreement in terms of PVC calls among the six programs we found a low score indicating that each program capture a specific profile for PVCs. Being the time of processing reasonable we suggest to explore the results of at least one filtering and one classifying method. We finally observed that a distinguishing feature in the most cited and applied RV packages VaxiJen and Vaxign, is their accessibility to final users through graphical user interfaces. We encourage researches in this field to invest in the development of user-friendly interfaces, as much as to the improvement of the predictive power of the algorithms.

AUTHOR CONTRIBUTIONS

All authors contributed to methods design, editing, and approved the final manuscript. AB and MD wrote and tested code, sourced data, performed data analysis, and drafted the manuscript. All authors read and approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.00113/full#supplementary-material>

REFERENCES

- Bowman BN, McAdam PR, Vivona S, Zhang JX, Luong T, Belew RK, et al. Improving reverse vaccinology with a machine learning approach. *Vaccine* (2011) 29:8156–64. doi: 10.1016/j.vaccine.2011.07.142
- Heinson AI, Gunawardana Y, Moesker B, Denman Hume CC, Vataga E, Hall Y, et al. Enhancing the biological relevance of machine learning classifiers for reverse vaccinology. *Int J Mol Sci.* (2017) 18:E312. doi: 10.3390/ijms18020312
- Rappuoli R. Reverse vaccinology. *Curr Opin Microbiol.* (2000) 3:445–50. doi: 10.1016/S1369-5274(00)00119-3
- Borrow R, Balmer P, Miller E. Meningococcal surrogates of protection - Serum bactericidal antibody activity. *Vaccine* (2005) 23:2222–7. doi: 10.1016/j.vaccine.2005.01.051
- Yang B, Sayers S, Xiang Z, He Y. Protegen: a web-based protective antigen database and analysis system. *Nucleic Acids Res.* (2011) 39:D1073–8. doi: 10.1093/nar/gkq944
- Ong E, Wong MU, He Y. Identification of new features from known bacterial protective vaccine antigens enhances rational vaccine design. *Front Immunol.* (2017) 8:1382. doi: 10.3389/fimmu.2017.01382
- Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci. USA.* (1988) 85:2444–8. doi: 10.1073/pnas.85.8.2444
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* (1990) 215:403–10. doi: 10.1016/S0022-2836(05)80360-2
- Nakai K, Horton P. PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci.* (1999) 24:34–5. doi: 10.1016/S0968-0004(98)01336-X
- Nielsen H, Engelbrecht J, Brunak S, Von Heijne G. A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Syst.* (1997) 8:581–99. doi: 10.1002/macp.201300342
- Hofmann K, Stoffel W. 103. Konferenz der gesellschaft für biologische chemie studiengruppe membranstruktur und -transport membranforum frankfurt. *Biol Chem.* (1993) 374:143–70. doi: 10.1515/bchm3.1993.374.1-6.143
- Serruto D, Bottomley MJ, Ram S, Giuliani MM, Rappuoli R. The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: immunological, functional and structural characterization of the antigens. *Vaccine* (2012) 30 (Suppl. 2):B87–97. doi: 10.1016/j.vaccine.2012.01.033
- Capo S, Nuti S, Scarselli M, Tavarini S, Montigiani S, Mori E, et al. Chlamydia pneumoniae genome sequence analysis and identification of HLA-A2-restricted CD8+ T cell epitopes recognized by infection-primed T cells. *Vaccine* (2005) 23:5028–37. doi: 10.1016/j.vaccine.2005.04.048
- Wizemann TM, Heinrichs JH, Adamou JE, Erwin AL, Kunsch C, Choi GH, et al. Use of a whole genome approach to identify vaccine molecules affording protection against *Streptococcus pneumoniae* infection. *Infect Immun.* (2001) 69:1593–8. doi: 10.1128/IAI.69.3.1593-1598.2001
- Ross BC, Czajkowski L, Hocking D, Margetts M, Webb E, Rothel L, et al. Identification of vaccine candidate antigens from a genomic analysis of *Porphyrromonas gingivalis*. *Vaccine* (2001) 19:4135–42. doi: 10.1016/S0264-410X(01)00173-6
- Masignani V, Rappuoli R, Pizzo M. Reverse vaccinology: a genome-based approach for vaccine development. *Expert Opin Biol Ther.* (2002) 2:895–905. doi: 10.1517/14712598.2.8.895
- Vivona S, Bernante F, Filippini F. NERVE: New Enhanced Reverse Vaccinology Environment. *BMC Biotechnol.* (2006) 6:35. doi: 10.1186/1472-6750-6-35
- Doytchinova IA, Flower DR. Vaxijen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* (2007) 8:4. doi: 10.1186/1471-2105-8-4
- He Y, Xiang Z, Mobley HLT. Vaxign: The first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *J Biomed Biotechnol.* (2010) 2010:297505 doi: 10.1155/2010/297505
- Jaiswal V, Chanumolu SK, Gupta A, Chauhan RS, Rout C. Jenner-predict server: Prediction of protein vaccine candidates (PVCs) in bacteria based on host-pathogen interactions. *BMC Bioinformatics* (2013) 14:211. doi: 10.1186/1471-2105-14-211
- Rizwan M, Naz A, Ahmad J, Naz K, Obaid A, Parveen T, et al. VacSol: A high throughput *in silico* pipeline to predict potential therapeutic targets in prokaryotic pathogens using subtractive reverse vaccinology. *BMC Bioinformatics* (2017) 18:106. doi: 10.1186/s12859-017-1540-0
- Goodswen SJ, Kennedy PJ, Ellis JT. Vacced: A high-throughput *in silico* vaccine candidate discovery pipeline for eukaryotic pathogens based on reverse vaccinology. *Bioinformatics* (2014) 30:2381–3. doi: 10.1093/bioinformatics/btu300
- Gardy JL, Laird MR, Chen F, Rey S, Walsh CJ, Ester M, et al. PSORTb v.2.0: Expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* (2005) 21:617–23. doi: 10.1093/bioinformatics/bti057
- Sachdeva G, Kumar K, Jain P, Ramachandran S. SPAAN: A software program for prediction of adhesins and adhesin-like proteins using neural networks. *Bioinformatics* (2005) 21:483–91. doi: 10.1093/bioinformatics/bti028
- Tusnányi GE, Simon I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* (2001) 17:849–50. doi: 10.1093/bioinformatics/17.9.849
- Brusic V, Rudy G, Harrison LC. MHCPEP, a database of MHC-binding peptides: update 1997. *Nucleic Acids Res.* (1998) 26:368–71. doi: 10.1093/nar/26.1.368
- Zaharieva N, Dimitrov I, Flower DR, Doytchinova I. Immunogenicity prediction by VaxiJen: a ten year overview. *J Proteomics Bioinformatics* (2017) 10:298–310. doi: 10.4172/jpb.1000454
- Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* (2003) 13:2178–89. doi: 10.1101/gr.1224503
- Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* (2010) 26:1608–15. doi: 10.1093/bioinformatics/btq249
- Bateman A. The Pfam protein families database. *Nucleic Acids Res.* (2004) 32:138D–141. doi: 10.1093/nar/gkh121
- Juncker AS, Willenbrock H, von Heijne G, Brunak S, Nielsen H, Krogh A. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* (2003) 12:1652–62. doi: 10.1110/ps.0303703
- Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Biocomputer* (2001) 2002:310–22. doi: 10.1142/9789812799623_0029
- Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* (2004) 4:1633–49. doi: 10.1002/pmic.200300771
- Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE* (2007) 2:e796. doi: 10.1371/journal.pone.0000796
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol.* (2000) 300:1005–16. doi: 10.1006/jmbi.2000.3903
- Cai R, Liu Z, Ren J, Ma C, Gao T, Zhou Y, et al. GPS-MBA: computational analysis of MHC class II epitopes in type 1 diabetes. *PLoS ONE* (2012) 7:e33884. doi: 10.1371/journal.pone.0033884
- Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. *Bioinformatics* (2009) 25:1293–9. doi: 10.1093/bioinformatics/btp137
- Duckert P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. *Protein Eng Des Sel.* (2004) 17:107–12. doi: 10.1093/protein/gzh013
- Kiemer L, Bendtsen JD, Blom N. NetAcet: prediction of N-terminal acetylation sites. *Bioinformatics* (2005) 21:1269–70. doi: 10.1093/bioinformatics/bti130
- Yu CS, Cheng CW, Su WC, Chang KC, Huang SW, Hwang JK, et al. CELLO2GO: A web server for protein subCELLular lOcalization prediction

- with functional gene ontology annotation. *PLoS ONE* (2014) 9:e99368. doi: 10.1371/journal.pone.0099368
41. Luo H, Lin Y, Gao F, Zhang CT, Zhang R. DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res.* (2014) 42:D574–80. doi: 10.1093/nar/gkt1131
 42. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* (2005) 33:D325–8. doi: 10.1093/nar/gki008
 43. UniProt Consortium T. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* (2018) 46:2699. doi: 10.1093/nar/gky092
 44. Kovacs-Simon A, Titball RW, Michell SL. Lipoproteins of bacterial pathogens. *Infect Immun.* (2011) 79:548–61. doi: 10.1128/IAI.00682-10
 45. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* (1960) 20:37–46. doi: 10.1177/001316446002000104

Conflict of Interest Statement: AB, DM, and RR were employees of GSK group of companies at the time of the study. MD is a Ph.D. student at the University of Turin and participates in a postgraduate studentship program at GSK.

Copyright © 2019 Dalsass, Brozzi, Medini and Rappuoli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Development of a Vaccine Against Meningococcus B Using Reverse Vaccinology

Vega Massignani¹, Mariagrazia Pizza¹ and E. Richard Moxon^{2*}

¹ GSK Vaccines, Siena, Italy, ² Department of Pediatrics, Oxford University, Oxford, United Kingdom

OPEN ACCESS

Edited by:

Lee Mark Wetzler,
Boston University, United States

Reviewed by:

Andrew Gorringe,
Public Health England,
United Kingdom
Scott D. Gray-Owen,
University of Toronto, Canada

*Correspondence:

E. Richard Moxon
richard.moxon@paediatrics.ox.ac.uk

Specialty section:

This article was submitted to
Vaccines and Molecular Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 06 December 2018

Accepted: 20 March 2019

Published: 16 April 2019

Citation:

Massignani V, Pizza M and Moxon ER
(2019) The Development of a Vaccine
Against Meningococcus B Using
Reverse Vaccinology.
Front. Immunol. 10:751.
doi: 10.3389/fimmu.2019.00751

The discovery of vaccine antigens through whole genome sequencing (WGS) contrasts with the classical hypothesis-driven laboratory-based analysis of microbes to identify components to elicit protective immunity. This radical change in scientific direction and action in vaccine research is captured in the term *reverse vaccinology*. The complete genome sequence of an isolate of *Neisseria meningitidis* serogroup B (MenB) was systematically analyzed to identify proteins predicted to be secreted or exported to the outer membrane. This identified hundreds of genes coding for potential surface-exposed antigens. These were amplified, cloned in expression vectors and used to immunize mice. Antisera against 350 recombinant antigens were obtained and analyzed in a panel of immunological assays from which 28 were selected as potentially protective based on the -antibody dependent, complement mediated- serum bactericidal activity assay. Testing of these candidate vaccine antigens, using a large globally representative strain collection of *Neisseria* species isolated from cases of disease and carriage, indicated that no single component would be sufficient to induce broad coverage and that a “universal” vaccine should contain multiple antigens. The final choice of antigens to be included was based on cross-protective ability, assayed by serum bactericidal activity and maximum coverage of the extensive antigenic variability of MenB strains. The resulting multivalent vaccine formulation selected consisted of three recombinant antigens (Neisserial Heparin Binding Antigen or NHBA, Factor H binding protein or fHbp and *Neisseria* Adhesin A or NadA). To improve immunogenicity and potential strain coverage, an outer membrane vesicle component obtained from the epidemic New Zealand strain (OMVNz) was added to the formulation to create a four component vaccine, called 4CMenB. A series of phase 2 and 3 clinical trials were conducted to evaluate safety and tolerability and to estimate the vaccine effectiveness of human immune responses at different ages and how these were affected by various factors including concomitant vaccine use and lot-to-lot consistency. 4CMenB was approved in Europe in 2013 and introduced in the National Immunization Program in the UK starting from September 2015 when the vaccine was offered to all newborns using a 2, 4, and 12 months schedule. The effectiveness against invasive MenB disease measured at 11 months after the study start and 5 months after the second vaccination was 83% and there have been no safety concerns.

Keywords: 4CMenB vaccine, reverse vaccinology, strain coverage, cross protection, antigenic variability, *Neisseria meningitidis* serogroup B (MenB)

MENINGOCOCCUS B: THE LAST FRONTIER

The development of a meningococcal vaccine to protect against invasive disease caused by serogroup B strains of *Neisseria meningitidis* (MenB) represents a milestone in vaccinology. MenB is a major cause of sepsis and meningitis in North and South America, Canada, Europe, Australasia, and many other countries, but developing an effective vaccine was for many years an unsolved problem. The stumbling block was that, in contrast to all other variant capsular polysaccharides of the meningococcus for which effective conjugate vaccines were developed and licensed, the B polysaccharide does not induce an effective antibody response. A study of 50 healthy adults immunized with MenB polysaccharide showed that all but three of them failed to produce any antibodies (1) and even conjugation to tetanus toxoid failed to improve its immunogenicity (2). In contrast, in the 1960s, it was shown that adult military recruits immunized with plain MenC polysaccharide responded with copious amounts of antibody that protected against meningitis and sepsis (3). Further, a trial in Finland using unconjugated MenA polysaccharide also showed strong, protective immune responses even in young children (4). Why was the MenB such a poor immunogen? Scientists from Finland showed that the MenB capsular polysaccharide was identical to sugars found on the surface of many human cells especially, but not exclusively, in the brain during its pre-natal development (5, 6). It was concluded that inducing antibodies to MenB capsular polysaccharide ran the risk of damaging structures found on the surface of human cells and the authors proposed that the immune system had evolved tolerance to the B polysaccharide as a mechanism to avoid autoimmune pathology because of mimicry between components on the cell surface of human cells and surface structures of bacteria. The B polysaccharide is a homo-polymer of α (2 \rightarrow 8) N-acetyl neuraminic acid, known also as polysialic acid (PSA), located on the surface of human cells. PSA has unusual and important biological properties. To allow intimate intercellular interactions, water must be excluded and PSA, richly hydrated, modulates the cell to cell signaling. Using antibodies or an enzyme that specifically destroys PSA, animal experiments have shown its key role in programming CNS development, including the migration of nerve cells, the connectivity between dendritic cells and the formation of junctions between muscle and nerves (7, 8).

Many scientists concluded that inducing antibodies to PSA in humans represented an unacceptable safety risk. A particularly alarming concern that consolidated opposition to using the polysaccharide as a vaccine concerned the risk to immunized women who become pregnant. Antibodies cross the placenta and reach the developing embryo, so antibodies resulting from immunization with B polysaccharide could disrupt CNS development in the unborn child, especially since the amount of PSA on neural tissues is known to be at its highest level during fetal development. But not all scientists were convinced that the evidence precluded using the B polysaccharide as a vaccine. Harold Jennings, one of the first to demonstrate that the B polysaccharide

was inert as an immunogen had the idea of modifying the B polysaccharide by introducing an N-propionyl side chain with the aim of increasing immunogenicity and precluding cross reactivity to human cells (9). This formulation elicited functional antibody responses in mice, but not in humans (10).

John Robbins was strongly supportive of this approach and even today remains steadfast in his opinion that using the B polysaccharide as a vaccine would not be harmful. He and his collaborators have documented that humans, through exposure to naturally occurring antigens, make antibodies to the polysaccharide, yet do not have an increased susceptibility to auto-immune disease (11). But, for the majority of scientists, an alternative approach to a vaccine, one that avoided the use of the potentially harmful B polysaccharide, was considered imperative. Crucially, vaccine manufacturers were strongly influenced by safety concerns and were unwilling to embark on investing millions into a research and development programs that risked being derailed when ethical approval was sought to carry out the mandatory clinical trials in humans.

Over many years, research on alternative approaches to develop a vaccine that protected against invasive diseases caused by strains of MenB was undertaken. Efforts were largely directed to the non-capsular antigens, proteins, or lipopolysaccharide (in meningococcus, often referred to as lipo-oligosaccharide or LOS). This change brought about a radical conceptual shift; referring to these non-capsule based vaccines as “group B” vaccines is a misnomer as they do not contain the defining feature of MenB bacteria (the polysaccharide capsule), and the vaccines will also potentially protect against other capsular groups (A, C, W, and Y strains).

The ability of the antibodies induced by each vaccine antigen to activate complement and induce bactericidal activity, measured in the serum bactericidal assay (SBA), was shown to be predictive of protection in humans (12, 13). However, although the SBA was well-established in the context of conjugate vaccines, its credentials as a correlate of protection in the context of protein antigens required further, independent validation. This came from the evidence of Goldschneider (12, 13) and experience with outer membrane vesicles (OMVs) [in particular those used in Norway and New Zealand (14, 15)], whose protective efficacy was shown to be largely mediated by bactericidal antibodies to the meningococcal surface protein, PorA. Although a number of surface-exposed candidate antigens were identified (16–18), none possessed sufficient capacity to elicit cross-reactive bactericidal activity against the diversity of meningococcal strains, the *sine qua non* that was essential for developing an effective vaccine. However, OMVs, treated with detergents to extract LOS and decrease endotoxin activity, were safe and effective in preventing group B meningococcal disease (19–22); A variety of “tailor-made” MenB OMV vaccines have been developed and licensed to control epidemics dominated by a single clone. OMV vaccines have been used in Norway (23), Cuba (24), Chile (25), and New Zealand (26). MeNZB, which was implemented in New Zealand, was associated with substantial reductions in invasive meningococcal disease caused by an outbreak clone

that reached an incidence of 17.4/100,000 of the population overall and more than 200/100,000 among some indigenous populations (27). In published studies, efficacy of two doses given to children 4 years or older, or to young adults, ranged from 57 to 83% but with the limitation that the protective, bactericidal responses of infants were specific for the major outer membrane porin protein (PorA) (14). Thus, the utility of OMV vaccines is limited to control clonal epidemics where disease is caused by strains expressing a PorA serosubtype matching that in the OMV vaccine. In an effort to broaden protection, OMV vaccines were prepared from several strains expressing distinct alleles of PorA, but the manufacturers of this multivalent formulation were not able to overcome a number of problems that included variable immunogenicity and consistency of formulation (28).

New technology was needed to overcome the impasse and this came about in 1995 when a team from The Institute for Genomic Research (TIGR) sequenced the complete genome of the human commensal-pathogen bacterium *Haemophilus influenzae* (29). The first completely assembled genome of a free-living organism was a revolution in biology; whole genome sequencing (WGS) transformed the scientific basis of epidemiology, diagnosis, and prevention of all diseases, including those caused by microbes. In the field of infectious diseases, WGS introduced a cost-effective method to acquire comprehensive information on pathogens and commensals, including those whose biology was elusive because they could not be cultivated in the laboratory. The implications of completing the first WGS were immediately apparent. The idea of using a genomics platform as a discovery tool to identify vaccine antigens was first explicitly published in 1997 (30) and the public health imperative to develop a vaccine to prevent invasive infections caused by MenB provided an ideal opportunity to exploit this concept. Scientists from Oxford University provided Venter and his team at The Institute for Genomics Research (TIGR) with DNA from a MenB strain (MC58) isolated from a UK outbreak of meningococcal disease (31). Preliminary genome sequence data, initially a modest 2-fold genome coverage, validated the potential of the approach by identifying a novel antigen (17, 18). Meantime, Italian scientists from Sclavo, Siena (led by Rino Rappuoli) had for several years dedicated their research efforts toward the development of meningococcal vaccines. In 1998, a collaboration between Chiron Vaccines (who acquired Sclavo), TIGR and Oxford University, carried out a comprehensive evaluation of all potential meningococcal vaccine antigens in the MC58 MenB strain, as described below. The WGS approach, highly sensitive but lacking the specificity to identify and prioritize antigens with respect to protective potential, was published in 2000 (32, 33). Antigen discovery through WGS contrasted with the hypothesis-driven classical laboratory based, bottom-up analysis of microbes to identify components that could elicit protective immunity. This radical change in scientific direction and action in vaccine research was subsequently captured in the term *reverse vaccinology* (34). Many of these were outer membrane proteins that had relatively low levels of surface expression, one reason why they had not been discovered before the use of WGS.

FROM BIOINFORMATICS TO BIOLOGY

Although the concept of mining genome information was straightforward, the challenge of whether it could result in the development of a vaccine was not. Indeed, proper cellular localization is the key attribute of a bacterial protein to be considered as a potential vaccine candidate. While proteins located in the cytosol are generally not good immunological targets, surface-associated structures are potentially accessible to the immune system and therefore more likely to induce a functional immune response. Based on this assumption, an *in silico* bioinformatics approach was used to identify novel antigens for vaccine development. The genome was therefore screened systematically to identify proteins predicted to be secreted or exported to the outer membrane, localized in the periplasm or in the inner membrane. Furthermore, selection was also extended to proteins containing amino acid signatures predictive of a possible role in the adhesion to host factors, as well as other virulence mechanisms. This was challenging since the MenB genome consisted of more than 2,000 predicted genes, only a minority of which coded for surface expressed molecules of potential utility as vaccine antigens (33). Although today sophisticated software and dedicated suites of programs exist to accurately predict a protein's cellular localization and potential biological function, this was not the case 20 years ago. In the late 1990s, interrogation of sequence data was in its infancy, the utility of many of the algorithms was not validated and annotations were often misleading. Management and interpretation of about two million base pairs of meningococcal genome sequence data were prone to errors. For instance, prediction of the start codons was based on the identification of the first ATG occurring after a previously identified stop codon. Unfortunately, this did not take into account either the presence of a correctly spaced Shine Dalgarno sequence, or the potential presence of less frequent start codons like TTG or GTG (coding for leucine or valine, respectively). For example, the annotation of GNA1870 (later renamed fHbp) was incorrect as a result of automatic procedures **Figure 1** and (35) is now one of the most important meningococcal antigens. The plethora of repetitive DNA elements in the genomes of meningococci was a deterrent to efficient assembly and unambiguous identification of genes because of frame shifts or sequencing errors. In the particular case of MC58 genome annotation, 65 open reading frames that contained stretches of repetitive DNA were identified. Of these, only 16 were previously known; the remainder was discovered through complete genome sequencing (36). The fact that some of the genes with DNA repeats encode for surface associated proteins poses a problem in terms of antigen selection, as phase variable genes, potentially generating escape variants, may not be ideal candidates for vaccine development. Despite these challenges, 18 months after the beginning of sequencing, 600 potential vaccine candidates were identified *in silico*. The highest proportion of these putative candidates was represented by integral membrane proteins (characterized by multiple hydrophobic domains), followed by periplasmic proteins, lipoproteins, and outer membrane and secreted proteins, the latter group representing <15% of the total. Interestingly, only



FIGURE 1 | Challenges in annotating correct start codons: the case of GNA1870. Nucleotide sequence of the intergenic region between *nmb1869* and *gna1870* genes in MenB strain MC58. The DNA sequence and deduced amino acid translation of the C-terminus of NMB1869 and N-terminal predicted lipoprotein signal peptide of GNA1870 are shown. The predicted start codons attributed automatically by softwares like Glimmer or GeneMark are indicated by a red box and a red circle, respectively. However, both ATG sites lacked correctly positioned ribosome binding site as well as -10 and -35 promoter elements. Most importantly, in both cases, amino acid translation would have resulted in an N-terminus of GNA1870 which lacked the features of a leader peptide, thus preventing the correct prediction of GNA1870 as a lipoprotein. Manual inspection identified the rare GTG start codon in position +1 as the real translation initiation codon. The putative ribosome binding site is shaded and two possible -10 and -35 promoter elements are indicated by dashed lines and dots, respectively.

half of the selected gene products displayed homologies to proteins of defined function, whereas the others had no clearly attributable functional role.

The genes coding for the 600 potential surface-exposed antigens were amplified from the genome of strain MC58 and cloned in expression vectors to generate Histidine (His) or GST (Glutathione S-transferase) tagged proteins. The fusion protein form, His or GST tagged, which showed higher solubility was purified from *E. coli* and used to immunize mice. The antisera raised against each recombinant antigen were analyzed in a panel of immunological assays: Western blot to confirm that the antigen was expressed in meningococcus and at the predicted molecular weight; flow cytometry by FACS (Fluorescence Activated Cell Sorter) to evaluate accessibility of the antigen to antibody binding on the meningococcal surface, and SBA to assess the ability of the antibodies to bind the antigen and to promote complement mediated bacterial killing. In addition, some of the antigens were tested for their ability to confer protection in the infant rat or mouse septicemia models by active or passive immunizations. Following this approach, 350 of the 600 predicted surface-exposed proteins were successfully expressed in *E. coli* and purified as recombinant proteins. The rate of expression was mainly driven by the intrinsic features of the selected antigens, with those containing more than one predicted transmembrane domain being the most difficult to express. Of the 350 candidate antigens, 91 proved to be surface-exposed, and 28 were able to elicit a bactericidal response. The identification of 28 new bactericidal antigens represented a real breakthrough in the field, considering that in more than 50 years of research only few bactericidal antigens were characterized (32) (Figure 2).

Selecting which of the most promising antigens should be included in the MenB vaccine required an approach informed by sequence analysis of individual surface expressed proteins to assess the extent of their variation within the natural population of meningococcal strains (associated with invasive disease and carriage) and by evaluation of the cross-bactericidal activity

of antisera raised against each of the antigens. A collection of strains was assembled by scientists at University of Oxford so as to be representative of the *Neisseria* species, based on multiple serogroups of *N. meningitidis*. Also included were strains of *Neisseria gonorrhoeae*, *Neisseria cinerea*, and *Neisseria lactamica* to evaluate the sequence conservation of the top cross-protective antigens. This sequence conservation analysis revealed a substantial degree of variability in at least some of the candidates, suggesting that no single component would be sufficient to induce broad coverage and that a “universal” vaccine should contain multiple antigens (Figure 2).

FORMULATING THE MULTICOMPONENT VACCINE AND FUNCTIONAL CHARACTERIZATION OF ITS COMPONENTS

The final choice of antigens to be included in this multivalent vaccine formulation was based on cross-protective ability, assayed by bactericidal activity and maximum coverage of the extensive antigenic variability of meningococcal B (MenB) strains. When initially identified, each candidate antigen was referred to as GNA (Genome derived *Neisseria* Antigen) followed by a number representing the position of the encoding gene in the genome. The three most promising antigens identified were GNA2132, GNA1870, and GNA1994. In subsequent studies, these three antigens were given names: NHBA, fHbp, and NadA, respectively, based on their functional activity (Figure 3).

NHBA: GNA2132 is a lipoprotein of around 420–490 amino acids and was shown to bind heparin and heparan-sulfate *in vitro*, through an Arginine-rich region. Because of this function, it has been named *Neisseria* heparin binding antigen (NHBA) (42). Binding to heparin affects survival of *Neisseria* in a human blood killing assay (42). The Arginine rich region plays also a key role in adhesion to eukaryotic cells (43). The *nhba* gene is

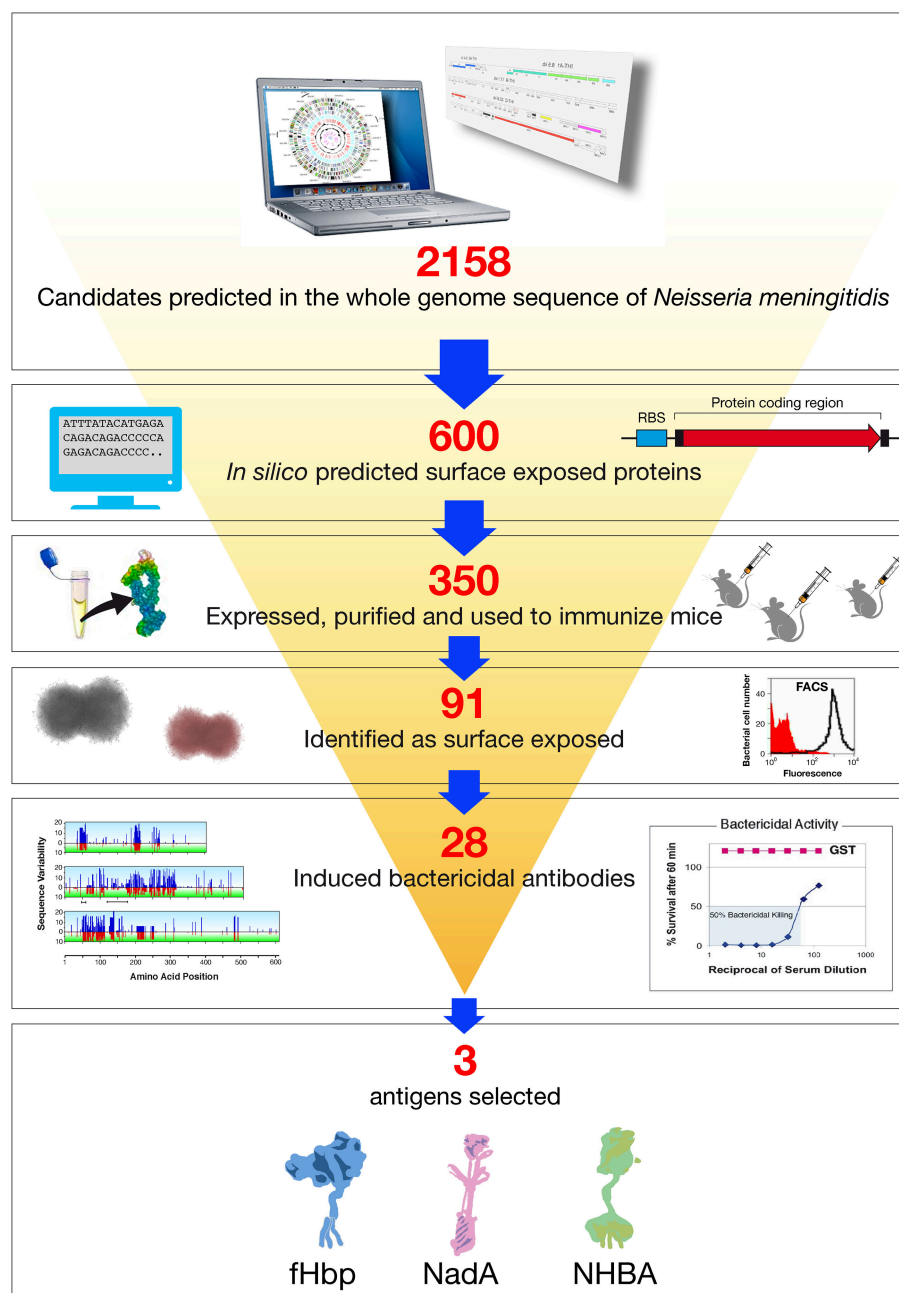
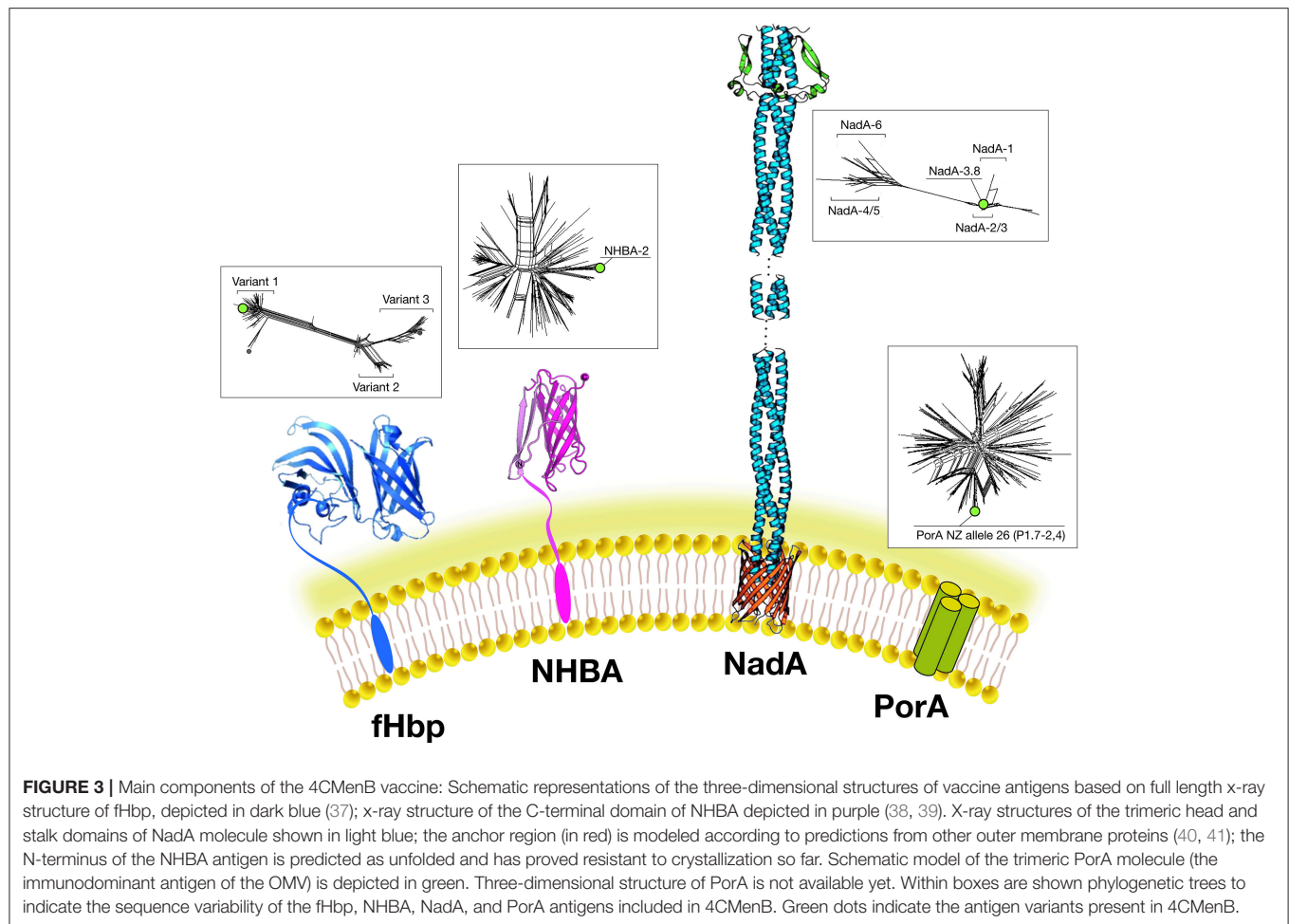


FIGURE 2 | Reverse vaccinology applied to *Neisseria meningitidis* B. Based on the complete genome sequence of *N. meningitidis* strain MC58, 2158 potential open reading frames (ORFs) were identified. *In silico* analysis resulted in the selection of 600 genes potentially encoding for novel surface exposed proteins. These were amplified by PCR and cloned into *Escherichia coli* expression vectors. Three hundred and fifty recombinant proteins were successfully produced, purified, and used to immunize mice. Recombinant protein candidates were then selected based on their surface expression (assessed by FACS), and ability to induce serum bactericidal antibodies (assessed by the SBA serum bactericidal assay) and conservation in a panel of *Neisseria* strains. The antigens selected by reverse vaccinology were finally prioritized, with NadA, fHbp, and NHBA as the three top antigens.

ubiquitous in all meningococcal serogroups and is also found in *N. gonorrhoeae* and other commensal *Neisseria* species. On the basis of the sequence variability, over 400 different peptides have been described and the relationship between sequence variability and cross-protection remains to be defined. The NHBA protein has an N-terminal region of approximately 250 residues

predicted to be intrinsically disordered, and a highly conserved C-terminal domain (~180 residues), with an 8-stranded anti-parallel β -barrel folding (38, 39). NHBA undergoes proteolytic cleavage by meningococcal NaIP protease and by eukaryotic proteases like human lactoferrin, Kallicrein, and the C3 convertase (42, 44).



fHbp: GNA 1870 is a lipoprotein of 253–266 amino acids able to bind human Factor H (FH), an inhibitor of the alternative complement pathway. Because of this activity it has been named factor H binding protein or fHbp (the same antigen discovered using a biochemical approach was named rLP2086) (35, 45). The binding of fHbp to FH enhances meningococcal serum resistance allowing the bacterium to replicate in human blood. The three-dimensional structures of fHbp alone or in complex with domains 6 and 7 of human FH have been solved. Interestingly, the side chains of fHbp that interact with FH resemble the glycosaminoglycan binding region of FH on host cells (37). Therefore, *Neisseria* is able, through fHbp, to recruit FH by mimicking the host. Sequence diversity analysis allowed identification of three variants, named variants 1, 2, and 3 (or subfamily A and B), serologically distinct and with only a low cross-bactericidal activity between variants 2 and 3 strains. The amount of fHbp expressed by different MenB strains is controlled by the fHbp promoter and can vary of at least 15-fold (46). FHbp contains multiple bactericidal epitopes and bactericidal activity of anti-fHbp antibodies varies according to the genetic diversity and level of expression of fHbp in the different strains (47).

NadA: GNA 1994 is a trimeric autotransporter belonging to the Oca family (oligomeric coiled-coil adhesins) of 323–405 amino acids. It mediates adhesion and invasion to epithelial cells and for this reason it has been named NadA (*Neisseria adhesin A*) (48). The *nadA* gene is not present in all meningococcal strains, and its presence is mainly associated with the hyperinvasive sequence type 8 (ST-8), ST-11, ST-32, and ST-213 clonal complexes (cc) but is rarely present in ST-41/44 and ST-269 cc isolates (49). Six variants exist of which NadA1, NadA2, and NadA3 are highly immunogenic and induce cross-reactive SBA responses. NadA4 is associated with carriage strains. NadA5 is rare and found in only a few invasive isolates (50). NadA expression levels vary among isolates and expression is upregulated by niche-specific signals via the transcriptional regulator NadR, which binds the NadA promoter and represses transcription. DNA-binding activity of NadR is attenuated by 4-hydroxyphenylacetic acid (4-HPA), a natural molecule released in human saliva, thus leading to the de-repression of *nadA* *in vivo* (51). Because of this tight regulation, the role of NadA in vaccine coverage may be underestimated *in vitro*. NadA forms stable trimers on the bacterial surface and mediates binding to epithelial cells through interaction with protein

receptor molecules differentially expressed by various epithelial cell lines. The three dimensional structure reveals a novel TAA (trimeric auto-transporter adhesins) organization made mostly of a coiled-coil with protruding wing-like structures forming a head-like domain (40).

With the aim of maximizing strain coverage while facilitating large-scale-manufacturing, fHbp, NHBA, and NadA were fused to additional candidate antigens, previously selected based on their ability to induce bactericidal activity and/or protection in animal models. More than 30 protein-protein fusions were generated and analyzed for their biochemical and immunological properties. Based on these analyses, GNA2132-GNA1030 and GNA2091-GNA1870 were the most stable and the most immunogenic in animal testing. Surprisingly, bactericidal activity induced by immunization with fHbp and NBHA was increased when each was fused to other antigens. In contrast, NadA was less immunogenic when fused to other antigens, probably because of the loss of its trimeric structure.

A vaccine consisting of three recombinant proteins, two protein-protein fusions plus a single antigen, named recombinant MenB vaccine (rMenB) was formulated with aluminum hydroxide and used to immunize mice. From a collection of 214 *N. meningitidis* clinical isolates (obtained from Europe, Canada, US, Australia, and New Zealand) to represent the global population diversity of invasive serogroup B isolates, bactericidal assays were performed on 85 strains using rabbit sera as exogenous complement source. The rMenB vaccine induced bactericidal antibodies against 78% of these strains. To improve immunogenicity and potential strain coverage of rMenB, an outer membrane vesicle component obtained from the epidemic New Zealand strain was added to the formulation to create a four component vaccine, called 4CMenB (52).

FROM WGS VACCINE ANTIGEN DISCOVERY TO A LICENSED VACCINE

Each component of the MenB vaccine had to satisfy a plethora of demanding regulatory conditions with respect to safety and immunogenicity, each having complex cost implications with respect to their manufacture and formulation. Six years pre-clinical research on toxicity, stability and immunogenicity were required before the two candidate MenB formulations, rMenB and 4CMenB, were approved for clinical trials that commenced

in adults in 2004 (53). Clinical trials of these vaccines in infants began in 2006 (54, 55), bypassing the conventional pathway involving a step-wise decrease in the age of subjects; primary school to pre-school to toddler to infant. This accelerated program in part reflected an awareness from the experience with OMV based vaccines that the breadth of immune response to meningococcal antigens was likely to be age dependent. Therefore, there was an imperative to evaluate the breadth of the immune response (i.e., cross-reactivity with non-vaccine variants of the vaccine antigens) in the age group most at risk of disease early in the vaccine's clinical testing.

A major challenge for these clinical trials was how to determine the immunogenicity of the vaccine candidates. In contrast to the previously licensed meningococcal glycoconjugate vaccines for which target antigens (distinct polysaccharide capsules) were invariant structures, the outer membrane proteins contained in rMenB and 4CMenB are variable in both primary sequence and level of expression. As in pre-clinical studies, this required judicious selection of MenB isolates on which to perform bactericidal assays, such that these distinct strains were representative of the diversity of invasive disease MenB target antigens. A further constraint was the small amount of serum that could be obtained in clinical trials involving infants; this limited the number of assays that could be performed to evaluate immunogenicity.

Thus, while the initial phase 1 human studies tested the immune response against 15 strains, the early phase 2 infant studies tested post-immunization sera against a panel of 7 strains (54, 55). Four of these strains were reference strains chosen to demonstrate the immunogenicity of individual vaccine antigens (fHbp, NHBA, NadA, and PorA) (Table 1). Immunization with 4CMenB induced bactericidal antibodies against a greater proportion of meningococcal strains than did rMenB. These findings were the basis of the decision to select 4CMenB, rather than rMenB, for further clinical development. A series of clinical trials evaluated how the immune response to the 4CMenB vaccine antigens was influenced by factors such as age of administration, concomitant vaccine use and lot-to-lot consistency.

4CMenB was approved in Europe in 2013 and introduced in the National Immunization Program in the UK starting from September 2015; the vaccine was offered to all newborns using a 2, 4, and 12 months schedule. The effectiveness against IMD measured at eleven months after the study and five

TABLE 1 | MenB indicator strains used to assess bactericidal activity induced by immunization with 4CMenB.

Classification of serogroup B meningococcal strains			Antigen gene sequence variant			
Designation	Sequence type	Clonal complex	PorA (VR2)	fHBP	NadA	NHBA
44/76-SL	32	32	P1.16	1.1	–	3
5/99	1349	8	P1.2	2.23	2	20
NZ98/254	42	41/44	P1.4	1.14	–	2
M10713	136	41/44	P1.16-3	2.24	–	10

H44/76 is the indicator strain for fHbp, 5/99 for NadA, NZ98/254 for PorA and M10713 for NHBA.

months after the second vaccination, was 82.9% (95% CI: 24.1 – 95.2). The wide confidence limits reflect the challenges of interpreting the post-implementation data in the short term, given the relatively small numbers of cases and the temporal fluctuations in rates of disease that are typical of meningococcal disease epidemiology. Nonetheless, following implementation of 4CMenB vaccine, the number of cases in vaccine-eligible infants was reduced by 50% (95% CI 36–71; $p = 0.0001$), compared to the pre-vaccine period. The long term impact of 4CMenB vaccine implementation on disease burden, disease severity and safety will continue as part of the National Surveillance program (56). There were also extensive phase 2 and 3 studies to investigate the safety and tolerability of the vaccine, of particular importance given the previous experience of the reactogenicity of OMV vaccines. These clinical trials, involving approximately 7,400 children under 11 years of age prior to licensure in Europe “EMA assessment report November 2012. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/002333/WC500137883.pdf”), demonstrated that ~60% of children receiving 4CMenB concomitantly with DTaP (Diphtheria, Tetanus, and acellular Pertussis) and pneumococcal conjugate vaccines experienced fever, compared to ~30% when these vaccines were given without 4CMenB. In infants, local and systemic reactions appeared to be more frequent when 4CMenB was co-administered with other vaccines, but medical attendance after vaccination and fever-related serious adverse events (SAEs) were rare. The occurrence of febrile seizures was comparable to that reported from other combination vaccine studies. Two cases occurred within 24 h after the first and another two cases after the second vaccination with 4CMenB and routine vaccines. These cases were assessed as possibly associated with vaccination but were deemed as mild and resolved spontaneously. Most other adverse events were common childhood illnesses or events consistent with solicited reactions and resolved at final follow up (57). After its introduction in the routine UK immunization program, there was an increase in presentations to Accident and Emergency and in hospital admissions for transient adverse events following immunization (58). In contrast, a suggestion of an association with Kawasaki disease in early clinical trials was not supported by post-implementation surveillance (59).

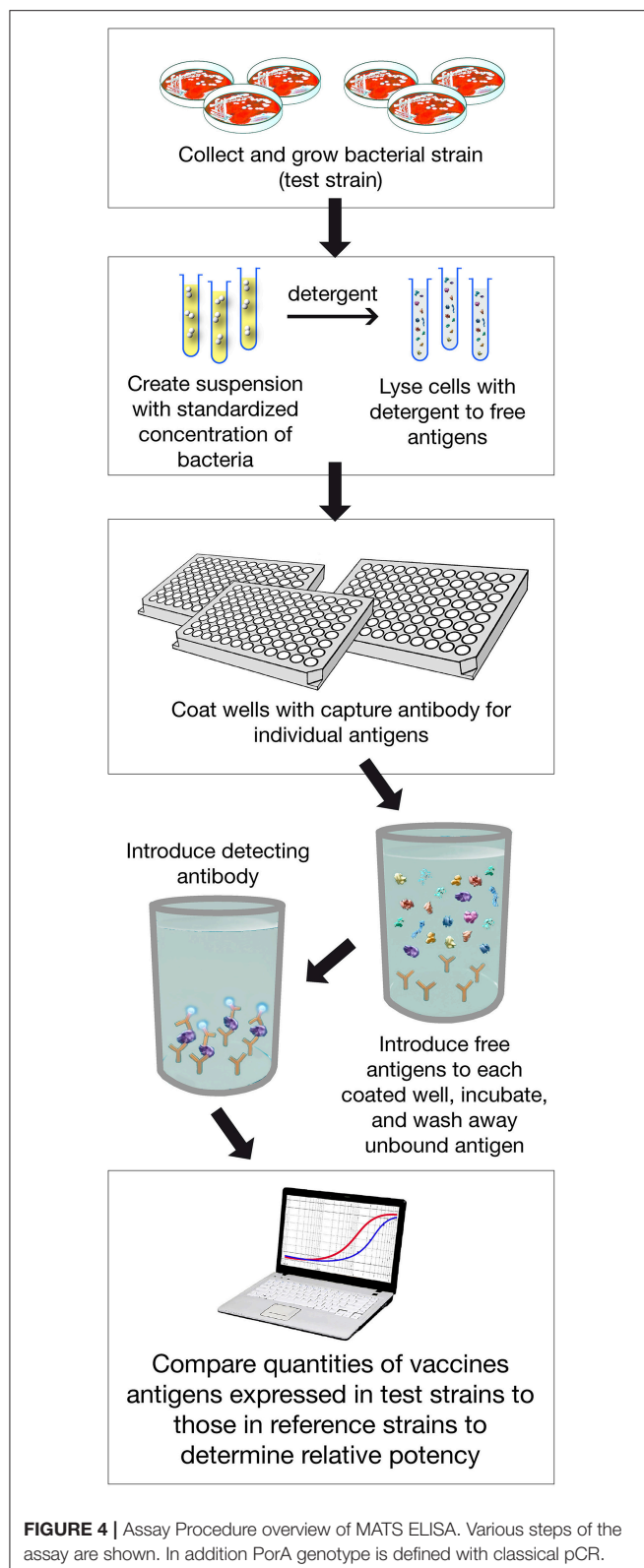
To obtain licensure by the European Medicines Agency and multiple other regulatory agencies internationally, a major challenge was how to estimate the protective potential of 4CMenB against invasive disease. Owing to the low incidence of meningococcal disease, classical efficacy studies were impractical. Thus, SBA using human complement was used to estimate vaccine functional immunity against invasive meningococcal disease. But, because the MenB strains that cause invasive meningococcal diseases are highly diverse with respect to the quantity and immunological cross-reactivity of the vaccine antigens expressed, estimating the effectiveness of the vaccine required performing SBA against large numbers of isolates, an undertaking that was judged to be impractical. Therefore, an innovative method was developed to assess coverage and predict effectiveness of the 4CMenB vaccine. This assay, called MATS (Meningococcal Antigen Typing System), correlated information

on the quantity and quality of the antigens expressed by individual MenB strains and the potency of the immune response elicited by the vaccine based on bactericidal assays.

MATS is based on the assumption that a given MenB strain is susceptible to killing by vaccine-induced antibodies, providing that this strain expresses one or more surface proteins in sufficient amounts so as to be adequately cross-reactive with a vaccine component (**Figure 4**). To develop the MATS assay, ELISA reactivity with antisera raised against fHbp, NHBA, and NadA expressed by each tested strain was compared to antigen specific reference MenB strains, a metric called relative potency (RP). Coverage of each individual strain is assumed if the RP is higher than an antigen-specific positive bactericidal threshold (PBT), defined for each antigen on the basis of bactericidal activity of infant sera against a panel of 57 serogroup B strains. PorA cross-reactivity is evaluated by exact sequence matching to PorA P1.4 vaccine serosubtype (60, 61). The MATS assay has been transferred to national reference laboratories in Europe, US and Australia. Worldwide, MATS-predicted coverage afforded by 4CMenB has been estimated at 66% in Canada (62) 68% in Portugal (63), 69% in Spain (64), 74% in Czech Republic (65), 76% in Australia (66), 78% in other 5 European countries (60), 81% in Brazil (66), 84% in Poland (67), 89% in Greece (68), and 91% in US (69). Finally, in England, Wales and Northern Ireland, the MATS estimate of coverage was measured at 73 and 67%, respectively, on different strain collections of 2007–2008 and 2014–2015 the latter representing the baseline before vaccine implementation (56) (**Figure 5**). Indeed, several publications (70, 71) claim a potential underestimation of MATS predicted coverage estimates, due to a series of reasons: (i) MATS provides an estimation of the contribution of each antigen independently, therefore the synergistic effect of antibodies recognizing different antigens is not measured; (ii) the NadA-mediated contribution to protection is underestimated as NadA expression is downregulated in the *in vitro* conditions in which MATS is performed, compared to expression of the antigen *in vivo* (51); (iii) the contribution of OMV to protection is limited to the presence of a matched PorA antigen, although it is commonly accepted that PorA-independent protection can be afforded by OMV against some strains.

The underestimation of protection predicted by MATS was further supported by a study performed in the UK on a panel of circulating clinical strains where the MATS predicted coverage was 73%, while the SBA showed 88% strain coverage (70). Similar data were also generated on a panel of MenB strains from Spain, showing that isolates found negative in MATS were in fact killed by sera of adolescents and infants immunized with 4CMenB (64). The overall underestimation has more recently been confirmed by the preliminary “real-world” effectiveness of 82.9% based on the results of the routine infant immunization with 4CMenB in the UK (56).

In Canada, 4CMenB was licensed in 2013 for use in 2 months to 18 years old. A mass vaccination campaign, targeting individuals aged 2 months to 20 years was implemented in the Saguenay-Lac-Saint-Jean region of Québec in 2014 to control the high incidence rate of MenB disease. Following the campaign, the incidence in the region decreased, with no



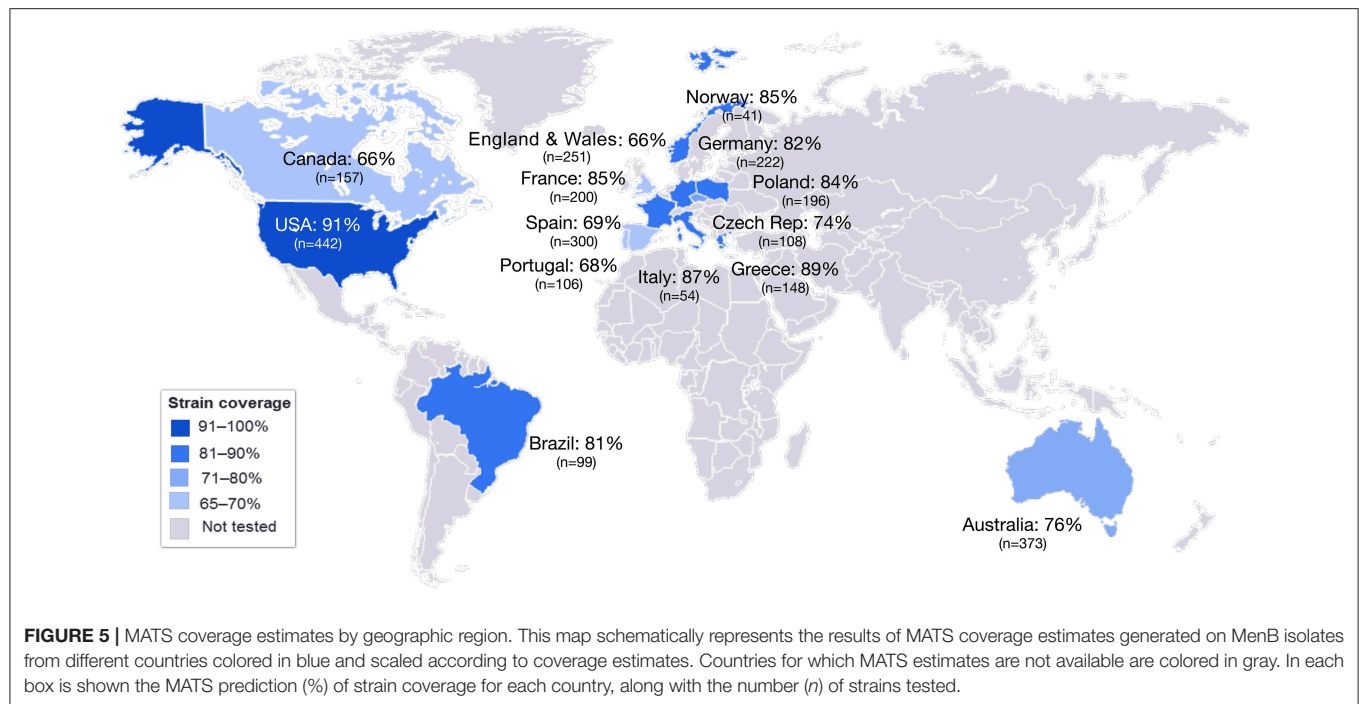
cases reported in the vaccinated individuals but with two cases occurring among the unvaccinated (72). In US, 4CMenB has been authorized in 2015 and recommended for use in the 10–25

years old as two doses vaccine. It has been used to control MenB outbreaks at University and college campus in Oregon, New Jersey and California (73, 74). No cases of MenB disease have been reported so far in vaccinated individuals, suggesting that the vaccine is effective in this age group. Moreover, when the immune responses induced by 4CMenB during the outbreak in Princeton was measured, 33% of 4CMenB vaccines showed no SBA against the outbreak strain, although no cases of meningococcal disease caused by *N. meningitidis* B were reported among vaccinated student (73).

DISCUSSION

The licensure in 2013 of the four component MenB vaccine (Bexsero) was the culmination of a scientific collaboration between university and industry-based scientists. The former provided cutting edge genomic, genetic, and clinical trials expertise; the latter undertook the vital high-throughput, “brute force” evaluation of hundreds of candidate antigens discovered through genomics, the in-depth characterization of the functional and immunological properties of the selected vaccine antigens and then stage-managed the pre-clinical and clinical testing required to obtain licensure. The facilitating technological breakthrough of WGS of bacterial pathogens came about through a former NIH academic, Craig Venter, who used his entrepreneurial vision to set up TIGR, the sequencing facility that made the MenB project possible. The 2018 Gairdner Award to Rino Rappuoli <https://www.aditecproject.eu/2017/05/04> who oversaw this academic-commercial partnership was fitting recognition of his role in driving through the innovative application of genomics to antigen discovery, the first example of what has become known as “reverse vaccinology.”

4CMenB represents a striking departure from the successful research and development platform that resulted in several, highly safe and effective conjugate meningococcal vaccines (against meningococcal serogroups A, C, W, and Y strains) formulated through covalent chemical linkage of different serogroup capsular polysaccharides to proteins. Although each of the meningococcal capsular polysaccharides shows strikingly distinct chemical compositions, each is an invariant structure whose target epitopes do not change over time or region. Diversity in the “carrier” proteins used to formulate conjugate vaccines are not problematic providing that these variations do not interfere with their role in recruiting T-cell help. But for vaccines, such as 4CMenB, where the antigens inducing protective immunity are proteins, the scenario is fundamentally different. The amino acid sequence of each of the protein antigens is highly variable, a consequence of their location on the bacterial surface where exposure to immune responses drives selection and fixation of diversity in circulating strains of meningococci. The multivalent protein vaccine, 4CMenB, is not without precedent; the several acellular pertussis vaccines have formulations consisting of up to 5 proteins, although in retrospect there was inadequate appreciation of the complications of allelic variation of these vaccine proteins. Loss or gain of DNA has



over many years impacted on the effectiveness of *B. pertussis* vaccines (75), but its population structure is clonal (76), so there is no recombination and the rate at which antigenic variation accumulates is very gradual over time. In contrast, genetic variation in meningococci occurs predominantly through recombination, not intra-genomic mutations. Thus, within the natural population of meningococci, there is frequent horizontal transfer of DNA, mainly through DNA transformation, not only between distinct genotypes of *N. meningitidis*, but also from other sub-species of *Neisseria* and, rarely, other distinct bacterial species. For example, conserved homologs of the *nhba* gene have been found in commensal *Neisseria* species, such as *N. lactamica*, *N. polysaccharaea*, and *N. flavescens* (77). This finding is relevant because of the potential “selective impact” that a NHBA-containing vaccine could have not only on encapsulated meningococcal strains, which are potentially pathogenic, but also on the commensal flora. This rampant recombination has major implications in that to be an effective vaccine, 4CMenB must elicit antibodies that protect against an enormous diversity of circulating meningococcal strains in a microbial population that is also constantly evolving over time. There was a requirement to develop a vaccine typing scheme to characterize any carriage or invasive meningococcal isolate. This effort resulted in the identification of the MATS assay as predictor of vaccine coverage. Since MATS can be applied only to cultivable strains, and considering that more than 50% of cases do not have an isolate, genomic driven predictor of coverage, such as BAST [Bexsero Antigen Sequence Typing (78)] or gMATS (genetic MATS), under development, will be instrumental to more precisely evaluate vaccine coverage.

The need for new lines of thinking emerged early in the pre-clinical phases of 4CMenB development. Given the rarity

of IMD [0.5–1 case per 100,000 per annum in Europe and the Americas] (79), reliance on a surrogate of protection to select appropriate protein antigens was paramount. The acceptance of SBA as a gold-standard surrogate of protection against meningococcal surface proteins (80) by scientists and regulatory authorities was a major milestone. It meant that the pre-clinical and clinical studies could proceed to licensure without the need for the conventional phase 3 efficacy trials for which cases of invasive disease provide the key metric. Given the logistics, expense, and large numbers of subjects required to assess efficacy, it was considered unlikely that any such clinical trial could be carried out.

Indeed, overall, the complexity of 4CMenB and the pathway to licensure made unprecedented demands on both the vaccine development teams and the regulatory authorities. Dialogue and an iterative scientific interchange was essential to address all regulatory requirements and translate them into practice.

The pathway of reliance on phase 2 immunogenicity studies of 4CMenB, backed by SBA and the derivative innovation of the MATS assay, was enormously facilitated by the previous experience with the MenC conjugate vaccines whose successful implementation in the routine UK immunization programme was a game changer (81). The way forward for 4CMenB has followed along similar lines, but has been immensely more complicated. For the reasons discussed above, estimates of effectiveness for the invariant meningococcal C polysaccharide vaccine antigen were far simpler than for the variable four protein antigens of 4CMenB. One major lesson emanating from experience with the conjugate vaccines in general, specifically exemplified by data on serogroup C meningococcal conjugates, has been the extent to which their success depends on indirect, or herd, immunity (82). The mechanism of indirect

protection is through curtailing transmission of meningococci and therefore decreasing the probability of new acquisitions and the risk of invasive disease. A UK study estimating the effect of meningococcal vaccines on herd protection against *N. meningitidis* in University students, showed that both, 4CMenB and MenACWY vaccines induced carriage reduction only for a subset of *Neisseria* strains, 4–12 months after vaccination (83). To date, the impact of 4CMenB on carriage of meningococci remains uncertain. Further studies are required, including those that ascertain whether there is an impact of the vaccine on bacterial load.

The inclusion of 4CMenB in the UK routine infant immunization programme since October 2014 allows post-implementation surveillance that over many years will provide crucial information on its effectiveness and duration of protection. In being given routinely only to infants, 4CMenB is not expected to prevent cases of meningococcal IMD in older children, adolescents or adults. In addition, estimates of protection, based on WGS, hSBA, and MATS, indicate that a proportion of strains lack a biologically relevant match to the antigens in the vaccine. As an example, the proportion of strains negative in hSBA in an UK strain panel was 12% (70). These *in silico* and *in vitro* predictions of vaccine effectiveness must be interpreted with caution since these metrics for estimating protection have not been validated. Two fundamentals of determining vaccine effectiveness are accurate information on immunization uptake and a robust system for disease notification. These data enable the calculation of vaccine effectiveness, as the likelihood of a child with disease being immunized (i.e., a vaccine failure) or unimmunized can be compared to that in the general population. This so-called screening method (84) was used to establish the effectiveness of the UK MenC conjugate vaccines and the OMV Vaccine in New

Zealand. It is crucial to have consensus on the definitions of what constitutes a case of meningococcal disease and an “immunized” child. To this end, organizations such as the European Centre for Disease Prevention and Control (ECDC) have provided definitions (85). More problematic is the definition of vaccine failure, which can be considered at either an individual or population level. At an individual level, not every case of MenB disease in an immunized child should be seen as a vaccine failure as the disease causing strain may not have expressed the vaccine target antigens. Thus, defining what criteria should be used to identify vaccine failures remains an exercise in pragmatism, dependent for validation on the accumulation of real-time data on rates of meningococcal disease, information that will require many years of surveillance using the screening method. Further complications include the changing incidence of IMD in the UK, a sharp decline in recent years (86) and the fact that in about half of all cases, no organism is isolated and confirmation is based on PCR (87), presenting major challenges to complete characterization of the target antigens of infecting meningococcal genotypes, in terms of expression and surface accessibility. 4CMenB is an exemplar of what can be truly considered a new era in vaccines.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

The authors thank Matthew D. Snape for his critical review of the manuscript and contributions to the clinical section.

REFERENCES

- Wyle FA, Artenstein MS, Brandt BL, Tramont EC, Kasper DL, Altieri PL, et al. Immunologic response of man to group B meningococcal polysaccharide vaccines. *J Infect Dis.* (1972) 126:514–21. doi: 10.1093/infdis/126.5.514
- Jennings HJ, Lugowski C. Immunochemistry of groups A, B, and C meningococcal polysaccharide-tetanus toxoid conjugates. *J Immunol.* (1981) 127:1011–8.
- Artenstein MS, Gold R, Zimmerly JG, Wyle FA, Schneider H, Harkins C. Prevention of meningococcal disease by group C polysaccharide vaccine. *N Engl J Med.* (1970) 282:417–20. doi: 10.1056/NEJM197002192820803
- Peltola H, Mäkelä H, Käyhty H, Jousimies H, Herva E, Hällström K, et al. Clinical efficacy of meningococcus group A capsular polysaccharide vaccine in children three months to five years of age. *N Engl J Med.* (1977) 297:686–91. doi: 10.1056/NEJM197709292971302
- Finne J, Bitter-Suermann D, Goridis C, Finne U. An IgG monoclonal antibody to group B meningococci cross-reacts with developmentally regulated polysialic acid units of glycoproteins in neural and extraneural tissues. *J Immunol.* (1987) 138:4402–7.
- Finne J, Leinonen M, Mäkelä PH. Antigenic similarities between brain components and bacteria causing meningitis. Implications for vaccine development and pathogenesis. *Lancet.* (1983) 2:355–7. doi: 10.1016/S0140-6736(83)90340-9
- Tomasiewicz H, Ono K, Yee D, Thompson C, Goridis C, Rutishauser U, et al. Genetic deletion of a neural cell adhesion molecule variant (N-CAM-180) produces distinct defects in the central nervous system. *Neuron.* (1993) 11:1163–74. doi: 10.1016/0896-6273(93)90228-J
- Rutishauser U. Influences of the neural cell adhesion molecule on axon growth and guidance. *J Neurosci Res.* (1985) 13:123–31. doi: 10.1002/jnr.490130109
- Jennings HJ, Gamian A, Ashton FE. N-propionylated group B meningococcal polysaccharide mimics a unique epitope on group B *Neisseria meningitidis*. *J Exp Med.* (1987) 165:1207–11. doi: 10.1084/jem.165.4.1207
- Bruge J, Bouveret-Le Cam N, Danve B, Rougon G, Schulz D. Clinical evaluation of a group B meningococcal N-propionylated polysaccharide conjugate vaccine in adult, male volunteers. *Vaccine.* (2004) 22:1087–96. doi: 10.1016/j.vaccine.2003.10.005
- Gottfredsson M, Reynisson IK, Ingvarsson RF, Kristjansdóttir H, Nardini MV, Sigurdsson JF, et al. Comparative long-term adverse effects elicited by invasive group B and C meningococcal infections. *Clin Infect Dis.* (2011) 53:e117–24. doi: 10.1093/cid/cir500
- Goldschneider I, Gotschlich EC, Artenstein MS. Human immunity to the meningococcus. II. Development of natural immunity. *J Exp Med.* (1969) 129:1327–48. doi: 10.1084/jem.129.6.1327
- Goldschneider I, Gotschlich EC, Artenstein MS. Human immunity to the meningococcus. I. The role of humoral antibodies. *J Exp Med.* (1969) 129:1307–26. doi: 10.1084/jem.129.6.1307
- Holst J, Oster P, Arnold R, Tatley MV, Næss LM, Aaberge IS, et al. Vaccines against meningococcal serogroup B disease containing outer membrane

- vesicles (OMV): lessons from past programs and implications for the future. *Hum Vaccin Immunother.* (2013) 9:1241–53. doi: 10.4161/hv.24129
15. Morley SL, Cole MJ, Ison CA, Camaraza MA, Sotolongo F, Anwar N, et al. Immunogenicity of a serogroup B meningococcal vaccine against multiple *Neisseria meningitidis* strains in infants. *Pediatr Infect Dis J.* (2001) 20:1054–61. doi: 10.1097/00006454-200111000-00010
 16. Granoff DM, Review of meningococcal group B vaccines. *Clin Infect Dis.* (2010) 50(Suppl. 2):S54–65. doi: 10.1086/648966
 17. Peak IR, Srikantha Y, Dieckelmann M, Moxon ER, Jennings MP. Identification and characterisation of a novel conserved outer membrane protein from *Neisseria meningitidis*. *FEMS Immunol Med Microbiol.* (2000) 28:329–34. doi: 10.1111/j.1574-695X.2000.tb01494.x
 18. Peak IR, Srikantha YN, Weynants VE, Feron C, Poolman JT, Jennings MP, et al. Evaluation of truncated NhhA protein as a candidate meningococcal vaccine antigen. *PLoS ONE.* (2013) 8:e72003. doi: 10.1371/journal.pone.0072003
 19. Frasch CE, van Alphen L, Holst J, Poolman JT, Rosenqvist E. Outer membrane protein vesicle vaccines for meningococcal disease. *Methods Mol Med.* (2001) 66:81–107. doi: 10.1385/1-59259-148-5:81
 20. Kelly C, Arnold R, Galloway Y, O'Hallahan J. A prospective study of the effectiveness of the New Zealand meningococcal B vaccine. *Am J Epidemiol.* (2007) 166:817–23. doi: 10.1093/aje/kwm147
 21. Bjune G, Grønnesby JK, Høiby EA, Closs O, Nøkleby H. Results of an efficacy trial with an outer membrane vesicle vaccine against systemic serogroup B meningococcal disease in Norway. *NIPH Ann.* (1991) 14:125–30.
 22. Noronha CP, Struchiner CJ, Halloran ME. Assessment of the direct effectiveness of BC meningococcal vaccine in Rio de Janeiro, Brazil: a case-control study. *Int J Epidemiol.* (1995) 24:1050–7. doi: 10.1093/ije/24.5.1050
 23. Fredriksen JH, Rosenqvist E, Wedege E, Bryn K, Bjune G, Frøholm LO, et al. Production, characterization and control of MenB-vaccine “Folkehelsa”: an outer membrane vesicle vaccine against group B meningococcal disease. *NIPH Ann.* (1991) 14:67–79.
 24. Sierra GV, Campa HC, Varcacel NM, Garcia IL, Izquierdo PL, Sotolongo PF, et al. Vaccine against group B *Neisseria meningitidis*: protection trial and mass vaccination results in Cuba. *NIPH Ann.* (1991) 14:195–207.
 25. Boslego J, Garcia J, Cruz C, Zollinger W, Brandt B, Ruiz S, et al. Efficacy, safety, and immunogenicity of a meningococcal group B (15:P1.3) outer membrane protein vaccine in Iquique, Chile. Chilean National Committee for Meningococcal Disease. *Vaccine.* (1995) 13:821–9. doi: 10.1016/0264-410X(94)00037-N
 26. Oster P, Lennon D, O'Hallahan J, Mulholland K, Reid S, Martin D. MeNZB: a safe and highly immunogenic tailor-made vaccine against the New Zealand *Neisseria meningitidis* serogroup B disease epidemic strain. *Vaccine.* (2005) 23:2191–6. doi: 10.1016/j.vaccine.2005.01.063
 27. Thornton V, Lennon D, Rasanathan K, O'Hallahan J, Oster P, Stewart J, et al. Safety and immunogenicity of New Zealand strain meningococcal serogroup B OMV vaccine in healthy adults: beginning of epidemic control. *Vaccine.* (2006) 24:1395–400. doi: 10.1016/j.vaccine.2005.09.043
 28. Cartwright K, Morris R, Rümke H, Fox A, Borrow R, Begg N, et al. Immunogenicity and reactogenicity in UK infants of a novel meningococcal vesicle vaccine containing multiple class 1 (PorA) outer membrane proteins. *Vaccine.* (1999) 17:2612–9. doi: 10.1016/S0264-410X(99)00044-4
 29. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science.* (1995) 269:496–512. doi: 10.1126/science.7542800
 30. Moxon ER. Applications of molecular microbiology to vaccinology. *Lancet.* (1997) 350:1240–4. doi: 10.1016/S0140-6736(97)03259-5
 31. Cartwright KA, Stuart JM, Noah ND. An outbreak of meningococcal disease in Gloucestershire. *Lancet.* (1986) 2:558–61. doi: 10.1016/S0140-6736(86)90124-8
 32. Pizza M, Scarlato V, Masignani V, Giuliani MM, Aricò B, Comanducci M, et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science.* (2000) 287:1816–20. doi: 10.1126/science.287.5459.1816
 33. Tettelin H, Saunders NJ, Heidelberg J, Jeffries AC, Nelson KE, Eisen JA, et al. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science.* (2000) 287:1809–15. doi: 10.1126/science.287.5459.1809
 34. Rappuoli R. Reverse vaccinology. *Curr Opin Microbiol.* (2000) 3:445–50. doi: 10.1016/S1369-5274(00)00119-3
 35. Masignani V, Comanducci M, Giuliani MM, Bambini S, Adu-Bobie J, Arico B, et al. Vaccination against *Neisseria meningitidis* using three variants of the lipoprotein GNA 1870. *J Exp Med.* (2003) 197:789–99. doi: 10.1084/jem.20021911
 36. Saunders NJ, Jeffries AC, Peden JE, Hood DW, Tettelin H, Rappuoli R, et al. Repeat-associated phase variable genes in the complete genome sequence of *Neisseria meningitidis* strain MC58. *Mol Microbiol.* (2000) 37:207–15. doi: 10.1046/j.1365-2958.2000.02000.x
 37. Schneider MC, Prosser BE, Caesar JJ, Kugelberg E, Li S, Zhang Q, et al. *Neisseria meningitidis* recruits factor H using protein mimicry of host carbohydrates. *Nature.* (2009) 458:890–3. doi: 10.1038/nature07769
 38. Esposito V, Musi V, de Chiara C, Veggi D, Scarselli M, et al. Structure of the C-terminal domain of *Neisseria* Heparin Binding Antigen (NHBA), one of the main antigens of a novel vaccine against *Neisseria meningitidis*. *J Biol Chem.* (2011) 286:41767–75. doi: 10.1074/jbc.M111.289314
 39. Maritan M, Veggi D, Cozzi R, Dello Iacono L, Bartolini E, Lo Surdo P, et al. Structures of NHBA elucidate a broadly conserved epitope identified by a vaccine induced antibody. *PLoS ONE.* (2018) 13:e0201922. doi: 10.1371/journal.pone.0201922
 40. Liguori A, Dello Iacono L, Maruggi G, Benucci B, Merola M, Lo Surdo P, et al. NadA3 structures reveal undecad coiled coils and LOX1 binding regions competed by meningococcus B vaccine-elicited human antibodies. *MBio.* (2018) 9. doi: 10.1128/mBio.01914-18
 41. Malito E, Biancucci M, Faleri A, Ferlenghi I, Scarselli M, Maruggi G, et al. Structure of the meningococcal vaccine antigen NadA and epitope mapping of a bactericidal antibody. *Proc Natl Acad Sci USA.* (2014) 111:17128–33. doi: 10.1073/pnas.1419686111
 42. Serruto D, Spadafina T, Ciucchi L, Lewis LA, Ram S, Tontini M, et al. *Neisseria meningitidis* GNA2132, a heparin-binding protein that induces protective immunity in humans. *Proc Natl Acad Sci USA.* (2010) 107:3770–5. doi: 10.1073/pnas.0915162107
 43. Vacca I, Del Tordello E, Gasperini G, Pezzicoli A, Di Fede M, Rossi Paccani S, et al. Neisserial Heparin Binding Antigen (NHBA) contributes to the adhesion of *Neisseria meningitidis* to human epithelial cells. *PLoS ONE.* (2016) 11:e0162878. doi: 10.1371/journal.pone.0162878
 44. Di Fede M, Biagini M, Cartocci E, Parillo C, Greco A, Martinelli M, et al. *Neisseria* Heparin Binding Antigen is targeted by the human alternative pathway C3-convertase. *PLoS ONE.* (2018) 13:e0194662. doi: 10.1371/journal.pone.0194662
 45. Mascioni A, Bentley BE, Camarda R, Dilts DA, Fink P, Gusarova V, et al. Structural basis for the immunogenic properties of the meningococcal vaccine candidate LP 2086. *J Biol Chem.* (2009) 284:8738–46. doi: 10.1074/jbc.M808831200
 46. Biagini M, Spinsanti M, De Angelis G, Tomei S, Ferlenghi I, Scarselli M, et al. Expression of factor H binding protein in meningococcal strains can vary at least 15-fold and is genetically determined. *Proc Natl Acad Sci USA.* (2016) 113:2714–9. doi: 10.1073/pnas.1521142113
 47. Seib KL, Scarselli M, Comanducci M, Toneatto D, Masignani V. *Neisseria meningitidis* factor H-binding protein fHbp: a key virulence factor and vaccine antigen. *Expert Rev Vaccines.* (2015) 14:841–59. doi: 10.1586/14760584.2015.1016915
 48. Capecchi B, Adu-Bobie J, Di Marcello F, Ciucchi L, Masignani V, Taddei A, et al. *Neisseria meningitidis* NadA is a new invasin which promotes bacterial adhesion to and penetration into human epithelial cells. *Mol Microbiol.* (2005) 55:687–98. doi: 10.1111/j.1365-2958.2004.04423.x
 49. Comanducci M, Bambini S, Brunelli B, Adu-Bobie J, Aricò B, Capecchi B, et al. NadA, a novel vaccine candidate of *Neisseria meningitidis*. *J Exp Med.* (2002) 195:1445–54. doi: 10.1084/jem.20020407
 50. Bambini S, De Chiara M, Muzzi A, Mora M, Lucidarme J, Brehony C, et al. *Neisseria* adhesin A variation and revised nomenclature scheme. *Clin Vaccine Immunol.* (2014) 21:966–71. doi: 10.1128/CI.00825-13
 51. Fagnocchi L, Biolchi A, Ferlicca F, Boccadifuoco G, Brunelli B, Brier S, et al. Transcriptional regulation of the nadA gene in *Neisseria meningitidis* impacts the prediction of coverage of a multicomponent meningococcal serogroup B vaccine. *Infect Immun.* (2013) 81:560–9. doi: 10.1128/IAI.01085-12

52. Giuliani MM, Adu-Bobie J, Comanducci M, Aricò B, Savino S, Santini L, et al. A universal vaccine for serogroup B meningococcus. *Proc Natl Acad Sci USA*. (2006) 103:10834–9. doi: 10.1073/pnas.0603940103
53. Toneatto D, Ismaili S, Ypma E, Vienken K, Oster P, Dull P. The first use of an investigational multicomponent meningococcal serogroup B vaccine (4CMenB) in humans. *Hum Vaccin*. (2011) 7:646–53. doi: 10.4161/hv.7.6.15482
54. Findlow J, Borrow R, Snape MD, Dawson T, Holland A, John TM, et al. Multicenter, open-label, randomized phase II controlled trial of an investigational recombinant meningococcal serogroup B vaccine with and without outer membrane vesicles, administered in infancy. *Clin Infect Dis*. (2010) 51:1127–37. doi: 10.1086/656741
55. Snape MD, Dawson T, Oster P, Evans A, John TM, Ohene-Kena B, et al. Immunogenicity of two investigational serogroup B meningococcal vaccines in the first year of life: a randomized comparative trial. *Pediatr Infect Dis J*. (2010) 29:e71–9. doi: 10.1097/INF.0b013e3181f59f6d
56. Parikh SR, Andrews NJ, Beebejaun K, Campbell H, Ribeiro S, Ward C, et al. Effectiveness and impact of a reduced infant schedule of 4CMenB vaccine against group B meningococcal disease in England: a national observational cohort study. *Lancet*. (2016) 388:2775–82. doi: 10.1016/S0140-6736(16)31921-3
57. Vesikari T, Esposito S, Prymula R, Ypma E, Kohl I, Toneatto D, et al. Immunogenicity and safety of an investigational multicomponent, recombinant, meningococcal serogroup B vaccine (4CMenB) administered concomitantly with routine infant and child vaccinations: results of two randomised trials. *Lancet*. (2013) 381:825–35. doi: 10.1016/S0140-6736(12)61961-8
58. Nainani V, Galal U, Buttery J, Snape MD. An increase in accident and emergency presentations for adverse events following immunisation after introduction of the group B meningococcal vaccine: an observational study. *Arch Dis Child*. (2017). doi: 10.1136/archdischild-2017-312941
59. Bryan P, Seabroke S, Wong J, Donegan K, Webb E, Goldsmith C, et al. Safety of multicomponent meningococcal group B vaccine (4CMenB) in routine infant immunisation in the UK: a prospective surveillance study. *Lancet Child Adolesc Health*. (2018) 2:395–403. doi: 10.1016/S2352-4642(18)30103-2
60. Vogel U, Taha MK, Vazquez JA, Findlow J, Claus H, Stefanelli P, et al. Predicted strain coverage of a meningococcal multicomponent vaccine (4CMenB) in Europe: a qualitative and quantitative assessment. *Lancet Infect Dis*. (2013) 13:416–25. doi: 10.1016/S1473-3099(13)70006-9
61. Donnelly J, Medini D, Boccadifuoco G, Biolchi A, Ward J, Frasch C, et al. Qualitative and quantitative assessment of meningococcal antigens to evaluate the potential strain coverage of protein-based vaccines. *Proc Natl Acad Sci USA*. (2010) 107:19490–5. doi: 10.1073/pnas.1013758107
62. Bettinger JA, Scheifele DW, Halperin SA, Vaudry W, Findlow J, Borrow R, et al. Diversity of Canadian meningococcal serogroup B isolates and estimated coverage by an investigational meningococcal serogroup B vaccine (4CMenB). *Vaccine*. (2013) 32:124–30. doi: 10.1016/j.vaccine.2013.03.063
63. Simões MJ, Bettencourt C, De Paola R, Giuliani M, Pizza M, Moschioni M, et al. Predicted strain coverage of a meningococcal multicomponent vaccine (4CMenB) in Portugal. *PLoS ONE*. (2017) 12:e0176177. doi: 10.1371/journal.pone.0176177
64. Abad R, Medina V, Stella M, Boccadifuoco G, Comanducci M, Bambini S, et al. Predicted strain coverage of a new meningococcal multicomponent vaccine (4CMenB) in Spain: analysis of the differences with other European countries. *PLoS ONE*. (2016) 11:e0150721. doi: 10.1371/journal.pone.0150721
65. Křižová P, Musilek M, Vacková Z, Kozáková J, Claus H, Vogel U, et al. Predicted strain coverage of a new protein-based meningococcal vaccine in the Czech Republic. *Epidemiol Mikrobiol Imunol*. (2014) 63:103–6.
66. Medini D, Stella M, Wassil J. MATS: global coverage estimates for 4CMenB, a novel multicomponent meningococcal B vaccine. *Vaccine*. (2015) 33:2629–36. doi: 10.1016/j.vaccine.2015.04.015
67. Waśko I, Hong E, De Paola R, Stella M, Moschioni M, Taha MK, et al. High predicted strain coverage by the multicomponent meningococcal serogroup B vaccine (4CMenB) in Poland. *Vaccine*. (2016) 34:510–15. doi: 10.1016/j.vaccine.2015.11.070
68. Tzanakaki G, Hong E, Kesanopoulos K, Xirogianni A, Bambini S, Orlandi L, et al. Diversity of Greek meningococcal serogroup B isolates and estimated coverage of the 4CMenB meningococcal vaccine. *BMC Microbiol*. (2014) 14:111. doi: 10.1186/1471-2180-14-111
69. Rajam G, Stella M, Kim E, Paulos S, Boccadifuoco G, Serino L, et al. Meningococcal Antigen Typing System (MATS)-based *Neisseria meningitidis* serogroup B coverage prediction for the MenB-4C vaccine in the United States. *mSphere*. (2017) 2:e00261-17. doi: 10.1128/mSphere.00261-17
70. Frosi G, Biolchi A, Lo Sapio M, Rigat F, Gilchrist S, Lucidarme J, et al. Bactericidal antibody against a representative epidemiological meningococcal serogroup B panel confirms that MATS underestimates 4CMenB vaccine strain coverage. *Vaccine*. (2013) 31:4968–74. doi: 10.1016/j.vaccine.2013.08.006
71. Abad R, Biolchi A, Moschioni M, Giuliani MM, Pizza M, Vázquez JA. A large portion of meningococcal antigen typing system-negative meningococcal strains from Spain is killed by sera from adolescents and infants immunized with 4CMenB. *Clin Vaccine Immunol*. (2015) 22:357–60. doi: 10.1128/CI.00669-14
72. De Wals P, Deceuninck G, Lefebvre B, Tsang R, Law D, De Serres G, et al. Impact of an immunization campaign to control an increased incidence of serogroup B meningococcal disease in one region of Quebec, Canada. *Clin Infect Dis*. (2017) 64:1263–7. doi: 10.1093/cid/cix154
73. Basta NE, Mahmoud AF, Borrow R. Meningococcal B vaccine during a university outbreak. *N Engl J Med*. (2016) 375:1595. doi: 10.1056/NEJMoa1514866
74. Grogan J, Roos K. Serogroup B meningococcus outbreaks, prevalence, and the case for standard vaccination. *Curr Infect Dis Rep*. (2017) 19:30. doi: 10.1007/s11908-017-0587-4
75. Poolman JT. Shortcomings of pertussis vaccines: why we need a third generation vaccine. *Expert Rev Vaccines*. (2014) 13:1159–62. doi: 10.1586/14760584.2014.944902
76. Bart MJ, Harris SR, Advani A, Arakawa Y, Bottero D, Bouchez V, et al. Global population structure and evolution of *Bordetella pertussis* and their relationship with vaccination. *MBio*. (2014) 5:e01074. doi: 10.1128/mBio.01074-14
77. Muzzi A, Mora M, Pizza M, Rappuoli R, Donati C. Conservation of meningococcal antigens in the genus *Neisseria*. *MBio*. (2013) 4:e00163-13. doi: 10.1128/mBio.00163-13
78. Brehony C, Rodrigues CMC, Borrow R, Smith A, Cunney R, Moxon ER, et al. Distribution of Bexsero(R) Antigen Sequence Types (BASTs) in invasive meningococcal disease isolates: implications for immunisation. *Vaccine*. (2016) 34:4690–7. doi: 10.1016/j.vaccine.2016.08.015
79. Rosenstein NE, Perkins BA, Stephens DS, Popovic T, Hughes JM. Meningococcal disease. *N Engl J Med*. (2001) 344:1378–88. doi: 10.1056/NEJM200105033441807
80. *The Immunological Basis for Immunization Series. Immunization, VaB Meningococcal Disease*. World Health Organization (2010) Module 15 (ISBN 978 92 4 15984 9).
81. Miller E, Salisbury D, Ramsay M. Planning, registration, and implementation of an immunisation campaign against meningococcal serogroup C disease in the UK: a success story. *Vaccine*. (2001) 20(Suppl. 1):S58–67. doi: 10.1016/S0264-410X(01)00299-7
82. Ramsay ME, Andrews NJ, Trotter CL, Kaczmarski EB, Miller E. Herd immunity from meningococcal serogroup C conjugate vaccination in England: database analysis. *BMJ*. (2003) 326:365–6. doi: 10.1136/bmj.326.7385.365
83. Read RC, Baxter D, Chadwick DR, Faust SN, Finn A, Gordon SB, et al. Effect of a quadrivalent meningococcal ACWY glycoconjugate or a serogroup B meningococcal vaccine on meningococcal carriage: an observer-blind, phase 3 randomised clinical trial. *Lancet*. (2014) 384:2123–31. doi: 10.1016/S0140-6736(14)60842-4
84. Farrington CP. Estimation of vaccine effectiveness using the screening method. *Int J Epidemiol*. (1993) 22:742–6. doi: 10.1093/ije/22.4.742
85. European Centre for Disease Prevention and Control. *Surveillance of Invasive Bacterial Diseases in Europe 2007*. (2010). Available online at: <https://ecdc.europa.eu>

86. Public Health England. *Invasive Meningococcal Disease in England: Annual Laboratory Confirmed Reports for Epidemiological Year 2016 to 2017*. Health protection Report (2017). p. 11.
87. Meningococcal Reference Unit, Gray SJ, Trotter CL, Ramsay ME, Guiver M, Fox AJ, et al. Epidemiology of meningococcal disease in England and Wales 1993/94 to 2003/04: contribution and experiences of the Meningococcal Reference Unit. *J Med Microbiol.* (2006) 55(Pt 7):887–96. doi: 10.1099/jmm.0.46288-0

Conflict of Interest Statement: VM and MP are employees of the GSK group of companies. EM has a consultancy contract with GSK.

Copyright © 2019 Masignani, Pizza and Moxon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Comprehensive Evaluation of the Expressed CD8+ T Cell Epitope Space Using High-Throughput Epitope Mapping

Paul V. Lehmann^{1*}, Maneewan Suwansaard¹, Ting Zhang¹, Diana R. Roen¹, Greg A. Kirchenbaum¹, Alexey Y. Karulin¹, Alexander Lehmann¹ and Pedro A. Reche²

¹ Cellular Technology Ltd., Shaker Heights, OH, United States, ² Laboratorio de Inmunomedicina & Inmunoinformática, Departamento de Immunología & O2, Facultad de Medicina, Universidad Complutense de Madrid, Madrid, Spain

OPEN ACCESS

Edited by:

Rene De Waal Malefyt,
Merck, United States

Reviewed by:

Katalin A. Wilkinson,
Francis Crick Institute,
United Kingdom
António Gil Castro,
University of Minho, Portugal

*Correspondence:

Paul V. Lehmann
paul.lehmann@immunospot.com

Specialty section:

This article was submitted to
T Cell Biology,
a section of the journal
Frontiers in Immunology

Received: 16 November 2018

Accepted: 11 March 2019

Published: 26 April 2019

Citation:

Lehmann PV, Suwansaard M, Zhang T, Roen DR, Kirchenbaum GA, Karulin AY, Lehmann A and Reche PA (2019) Comprehensive Evaluation of the Expressed CD8+ T Cell Epitope Space Using High-Throughput Epitope Mapping. *Front. Immunol.* 10:655. doi: 10.3389/fimmu.2019.00655

T cell immunity is traditionally assessed through functional recall assays, which detect the consequences of the T cells' antigen encounter, or via fluorescently labeled multimers that selectively bind peptide-specific T cell receptors. Using either approach, if the wrong antigen or peptide of a complex antigenic system, such as a virus, is used for immune monitoring, either false negative data will be obtained, or the magnitude of the antigen-specific T cell compartment will go largely underestimated. In this work, we show how selection of the "right" antigen or antigenic peptides is critical for successful T cell immune monitoring against human cytomegalovirus (HCMV). Specifically, we demonstrate that individual HCMV antigens, along with previously reported epitopes, frequently failed to detect CD8+ T cell immunity in test subjects. Through systematic assessment of T cell reactivity against individual nonamer peptides derived from the HCMVpp65 protein, our data clearly establish that (i) systematic testing against all potential epitopes encoded by the genome of the antigen of interest is required to reliably detect CD8+ T cell immunity, and (ii) genome-wide, large scale systematic testing of peptides has become feasible through high-throughput ELISPOT-based "brute force" epitope mapping.

Keywords: epitope, peptide, HCMV, EBV, MHC, HLA, T cell, ELISPOT

INTRODUCTION

Unlike B cells that endow humoral immunity through secretion of immunoglobulins, T cells exert their protective functions through direct interactions with antigen bearing cells. Conventional T cells recognize, via their T cell receptor (TCR), processed peptide fragments derived from polypeptide antigens, hereafter referred to as epitopes. These epitopes are displayed for T cell recognition on dedicated antigen-presenting molecules that are encoded by the major histocompatibility complex (MHC), which in humans is termed human leukocyte antigen complex (HLA). The TCR binds a binary complex of the epitope aligned in the peptide-binding groove of the MHC molecule (1, 2) displayed on the surface of antigen presenting cells (APC).

The TCR of the CD8+ subpopulation of T cells (CD8+ T cells) recognize epitopes presented in the context of MHC Class I molecules (MHC-I), while the TCR of CD4+ T cells recognize epitopes presented by Class II (MHC-II) molecules. During infection or vaccination, MHC molecules on

APC will be loaded with peptide fragments derived from the individual proteins (antigens) that constitute the infectious agent (the antigen-system), and are transported to the APC's surface for T cell recognition (3). The specific peptide fragments from an antigen that are presented to T cells is primarily dictated by the MHC molecules encoded by an individual, and like the MHC molecules themselves, underlies considerable inter-individual variability. MHC I and II molecules are polygenic, and moreover each locus underlies extensive allelic polymorphism. Consequently, individuals in an outbred population will express a unique set of MHC molecules (4). As the polygenism and polymorphism of MHC primarily affects the peptide binding groove, each MHC locus and allele essentially encode for a distinct peptide binding motif, i.e., peptide binding specificity (5). While human subjects can share a common HLA-allele, there are few, if any, unrelated individuals on this planet who express the same constellation of HLA molecules. For this reason, it is extremely unlikely that two humans on this earth whose T cell system, when faced with the need to recognize a given infectious organism, would recognize an identical set of epitopes. The *potential epitope space*, that is, the array of peptides that meet MHC binding criteria in a given host is, therefore, unique to that individual. The highly individualized MHC-restricted recognition of epitopes by T cells is thought to have evolved to protect the species against antigenic mimicry: if a pathogen succeeds to evade immune recognition in one individual, or a subset of individuals sharing a certain MHC allele, it might endanger this subject or subpopulation, but not the species as a whole (6).

The T cell response elicited by an infection or immunization is dictated not only by the above inter-individual diversity of epitopes displayed for T cell recognition, but also by the inherent variability in the TCR repertoire possessed by different individuals. An individual's TCR repertoire is shaped by genetics, but also through selection during the development of self-tolerance, and by foreign antigens during environmental exposures or infections (7, 8). Due to such TCR repertoire variations, different individuals' T cells will not respond uniformly to the *potential epitope space*. For example, even if individuals share an HLA molecule, such as the HLA-A*02:01 allele, which should present the same set of peptides, these individuals will not generate a uniform T cell response to these peptides (9). The array of epitopes actually recognized by T cells (constituting the *expressed T cell repertoire*) can therefore differ between individuals even if the potential epitope space (dictated by the MHC) is identical or overlapping between these individuals. The expressed T cell repertoire encompasses only a fraction of the expressed epitope space (10), but being dictated by the latter, will also be highly unique to each individual.

T cell immune monitoring aims at defining the expressed T cell repertoire, that is, the magnitude and quality of the antigen-specific T cells that have been generated in the course of the immune response. The magnitude of the expressed T cell repertoire is determined by the sum of all T cells that target all individual epitopes of the antigen system (e.g., virus) that have triggered a T cell response in the individual. Likewise, the quality of the elicited T cell response is defined by the effector functions of these T cells, such as the type of cytokine they secrete, or

whether they are cytolytic. Given the multitude and variability of epitopes (peptides) of an antigenic system recognized in different individuals, a major challenge for T cell monitoring has been identification of the "relevant" epitopes (i.e., test peptides) for each test subject. In an attempt to address this, the mainstream approach has focused on one, or very few, epitopes restricted by an individual HLA allele and to study T cell recognition of peptides that have either been previously reported, or predicted, to be restricted by that allele due to its peptide binding motif (11). Owing to the high prevalence of the HLA-A*02:01 allele in the Caucasian population, immune monitoring has frequently focused on peptides restricted to this particular HLA-A allele. Approximately half of the Caucasian population expresses the HLA-A*02:01 allele (12), but it is less frequent in other races.

Although narrowing immune monitoring to single HLA-alleles is common, presently it is unclear how adequate this approach is. Any T cell immune monitoring effort that restricts itself to one or few epitopes presented by one HLA allele of a test subject inherently has the disadvantage of neglecting the remainder of HLA alleles expressed by the test subject. For example, selecting a single HLA Class I allele (one of up to six capable of presenting epitopes to CD8+ T cells in each test subject), and one test peptide capable of binding to the selected HLA molecule, can only be a valid approach to CD8+ T cell immune monitoring if CD8+ T cells indeed prevalently target that single epitope in the test individuals. Such immune dominance, however, might be the exception in humans, as we tend not only to respond to a multitude of epitopes of an antigen, but also with highly variable clonal sizes to the individual epitopes (9). Therefore, ideally, T cell immune monitoring should comprehensively include the entire potential epitope space, encompassing all peptides that could be restricted by all HLA alleles present in every test subject. This can be accomplished by using pools of such peptides, or by testing all peptides individually, in a "brute force" high-throughput approach (13–15). The former, while much simpler in scope, enables comprehensive assessment of the expressed antigen-specific T cell repertoire, but does not reveal the epitope specificity of the antigen-reactive T cells, whereas the latter also provides this valuable information.

The ability to perform high-throughput testing of hundreds, even tens of thousands of peptides, on individual test subjects has only recently become technologically feasible (16, 17). The hurdles that needed to be overcome included limitations in peripheral blood mononuclear cell (PBMC) numbers available from humans, ease of access to extensive custom peptide libraries, high-throughput T cell assay platforms, and automated data analysis including databasing. While such brute force epitope mapping seems to call for a major effort that can be undertaken only by industrial scale laboratories with essentially unlimited manpower and budgets, we have developed and report here high-throughput epitope mapping strategies that can be readily applied in even small academic laboratories operating on tight budgets. In this paper we outline how it can be implemented. Before doing so, using CD8+ T cell immunity to HCMV (and, **Supplementary Material** for EBV and influenza virus) as a model, we illustrate how incomplete our understanding of an

individual's epitope space is to date, and thus how far we still are from comprehensive T cell immune monitoring. We also show the feasibility of systematic high-throughput ("brute force") epitope scans as the necessary next step to fill this void.

MATERIALS AND METHODS

Peripheral blood mononuclear cells (PBMC) from healthy human donors (HD) were obtained from CTLs ePBMC library (CTL, Shaker Heights, OH, USA). The PBMC had been collected by HemaCare Blood Donor Center (Van Nuys, CA) under HemaCare's IRB and sold to CTL identifying donors by code only while concealing the subjects' identity. All PBMC were from healthy adults who had not taken medication within a month of the blood draw that might influence their T cell response. In addition, tests were done on each donor at HemaCare's CLIA certified laboratory to identify common infections: serological testing was done for Syphilis, Human cytomegalovirus (HCMV), Epstein Barr Virus (EBV), Hepatitis B, Hepatitis C, Human Immunodeficiency Virus, Human T-Lymphotropic Virus, and *Trypanosoma cruzi*. Subjects positive for any of these infections, except for HCMV and EBV were disqualified for the ePBMC library. The CMV and EBV serological status is specified for donors and cohorts in the manuscript. The donor's age, sex, ethnicity and HLA-type is shown in **Supplementary Table 1**. HLA-typing was done by the HLA Typing Laboratory of University of Oklahoma Health science Center (Oklahoma City, OK). For all experiments donor cohorts were selected according to the HCMV or EBV serostatus specified, and then tested in T cell assays with the T cell assay results shown. The cryopreserved cells were thawed following an optimized protocol (18) resulting in viability exceeding 90% for all samples. The PBMC were resuspended in CTL-Test™ Medium (from CTL). CTL-Test™ Medium is serum-free and has been developed for low background and high signal performance in ELISPOT assays. The number of PBMC plated in the different ELISPOT experiments varied between 2 and 4×10^5 PBMC per well, with the number specified in Results for each experiment. The ELISPOT data are expressed as spot forming cells per million input PBMC (SFC/ 1×10^6 PBMC).

Antigens and Peptides

The peptide pools representing the HCMV antigens listed in **Table 1**, and EBV antigens specified in **Supplementary Table 2**, consisted of 15-mer peptides that covered the entire amino acid (aa) sequence of the respective proteins in steps (gaps of) 11 aa. Each of these peptide pools was purchased from JPT (Berlin, Germany), and were tested at a final concentration of 1 µg/mL in ELISPOT assays. The HCMV pp65_{495–503} peptide is an immunodominant peptide of HCMV that is presented by the HLA-A*02:01 allele (19). This HCMV pp65_{495–503} peptide, along with additional HCMV, EBV, and influenza virus peptides specified in **Figure 3**, **Supplementary Figures 2, 3**, respectively, were purchased from Panatecs (Heilbronn, Germany) at >95% purity according to the manufacturer's specification. Each of these purified, single peptides were tested at 1 µg/mL in ELISPOT

assays. HCMV and EBV grade 2 antigens (UV-inactivated virions) were purchased from Microbix (Mississauga, Ontario, Canada) and tested at a final concentration of 30 µg/mL. CPI (from CTL) was used as a positive control in all experiments because unlike CEF peptides, CPI elicits T cell recall responses in all healthy donors (20). CEF is a pool of 32 immune dominant nonamer peptides derived from CMV, EBV and the influenza virus commonly used as a positive control for CD8+ T cell recall (20). CPI is a combination of protein antigens derived from CMV, influenza and parainfluenza viruses, and was used at a final concentration of 6 µg/mL in ELISPOT assays.

The 553 nonamer peptides, spanning the entire HCMV pp65 sequence in steps of single aa were purchased from JPT as a FastTrack CD8 epitope library, and are hereafter referred to as pp65 9-mer peptides. Individual pp65 9-mer peptides were not subjected to further purification following their synthesis; however, individual peptides were verified and quantified by JPT using LC-MS. The average purity of the pp65 9-mer peptides was 56%, and the purity of individual 9-mer peptides that elicited CD8+ T cell recall responses are specified in **Table 2** for each respective peptide. The pp65 9-mer peptides were delivered as lyophilized powder with ~25 µg of a single 9-mer peptide present in a designated well of a 96-well plate, and were distributed across six 96-well plates. Individual peptides were initially dissolved using 50 µL DMSO, followed by addition of 200 µL of CTL-Test Medium generating a "primary peptide stock solution" at 100 µg/mL with 20% v/v DMSO. From each of these wells, using a 96-well multichannel pipettor, a "secondary, 10X, peptide stock solution" was prepared, with peptides at 2 µg/mL. On the day of testing, 20 µL from each well was transferred "en block," with a 96-well multi-channel pipettor into pre-coated ImmunoSpot® assay plates containing 80 µL CTL-Test Medium. Finally, 100 µL of PBMC (containing 3×10^5 cells) in CTL-Test media was added "en block" to achieve a final peptide concentration of 0.2 µg/mL in the ELISPOT assay.

Human IFN-γ ELISPOT Assays

Single-color enzymatic ImmunoSpot® kits from CTL were used for the detection of IFN-γ producing cells. Test procedures followed the manufacturer's recommendations. In brief, antigens/peptides were plated at the specified concentrations into capture antibody-precoated ELISPOT assay plates in a volume of 100 µL per well, dissolved in CTL-Test Medium. The plates with the antigen were stored at 37°C in a CO₂ incubator for less than an hour until the PBMC were ready for plating. The PBMC were added in the numbers specified (between 200,000 and 400,000 cells/well) in 100 µL CTL-Test Medium and cultured with the antigens/peptides for 24 h at 37°C and 9% CO₂ in an incubator. After removal of the cells, addition of detection antibody, and enzymatic visualization of plate-bound cytokine, the plates were air-dried prior to scanning and counting of spot forming units (SFU). ELISPOT plates were analyzed using an ImmunoSpot® EpiScan™ Reader, by CTL. SFU were automatically calculated by the ImmunoSpot® Software for each stimulation condition using the Autogate™ function (15).

TABLE 1 | HCMV-seropositive and -seronegative healthy human donors' reactivity to 20 HCMV antigens represented by peptide pools.

	Donor ID#	HCMV-seropositive donors						HCMV-seronegative donors					
		64	99	137	182	193	194	7	65	141	151	158	218
HCMV antigens (No. of peptides per pool)	Medium	28	5	20	48	0	3	8	5	8	3	18	0
	IE-1 (120) ^a	88	116	1,380	2,416	696	988	0	0	12	0	0	0
	IE-2 (143) ^b	16	72	52	92	244	24	36	0	0	0	4	0
	pp65 (138) ^c	532	1,412	1,272	216	1,092	340	0	4	0	0	12	4
	UL28 (92) ^d	8	228	16	684	32	520	0	0	0	0	0	4
	UL32 (260) ^e	16	2,320	504	80	380	100	8	0	0	0	0	4
	UL36 (117) ^f	17,772	192	24	936	348	124	4	0	0	4	0	0
	UL40 (53) ^g	0	4	24	76	20	8	0	0	0	0	24	4
	UL48-sub1 (229) ^h	8	8	16	36	4	20	0	0	0	0	0	0
	UL48-sub2 (229) ⁱ	100	16	36	112	24	4	4	0	4	0	8	0
	UL55 (224) ^j	2,340	104	2,308	2,328	1,628	48	4	0	4	0	4	0
	UL82(137) ^k	0	132	128	472	4	4	0	0	4	0	0	0
	UL94(84) ^l	148	0	52	2,492	724	20	0	12	0	0	12	4
	UL99(45) ^m	0	0	152	68	16	24	0	0	0	4	0	0
	UL103(60) ⁿ	96	96	52	80	4	24	4	0	0	0	4	0
	UL151(82) ^o	0	0	56	36	0	4	0	0	0	0	0	0
	UL153(67) ^p	8	36	136	68	24	20	0	0	0	0	4	0
	US3(44) ^q	12	0	48	1,152	12	12	0	0	8	0	0	0
	US24(123) ^r	0	1,468	68	2,172	672	92	8	0	0	0	0	0
	US29(113) ^s	4	0	40	64	0	12	0	0	0	0	12	0
	US32(43) ^t	4	0	128	104	12	24	0	0	4	0	0	0
	CPI	1,203	2,073	608	1,738	1,800	655	443	843	805	553	1,038	665

Pools of 15-mer peptides were tested that span the polypeptide sequence of each specified HCMV protein in steps of 11 amino acids, with the number of peptides contained within each pool specified in parentheses. Media was used as negative control and CPI as the positive control. The results are expressed as SFU/ 1×10^6 PBMC.

^aImmediate-Early protein 1 of HCMV; ^bImmediate-Early protein 2 of HCMV; ^c65 kDa phosphoprotein (pp65) of HCMV; ^dUncharacterized protein UL28 of HCMV; ^eLarge structural phosphoprotein UL32 of HCMV; ^fUncharacterized protein UL36 of HCMV; ^gUncharacterized protein UL40 of HCMV; ^hsubpool of Deneddylase UL48 of HCMV; ⁱsubpool of Deneddylase UL48 of HCMV; ^jEnvelope glycoprotein B (UL55) of HCMV; ^k71 kDa phosphoprotein (UL82) of HCMV; ^lCapsid-binding protein UL94 of HCMV; ^mTegument protein UL99 of HCMV; ⁿProtein UL103 of HCMV; ^oOrf UL151 of HCMV; ^pOrf UL153 of HCMV; ^qUnique short US3 glycoprotein of HCMV; ^rTegument protein US24 of HCMV; ^sUncharacterized protein HHRF4 (US29) of HCMV; ^tUncharacterized protein HHRF7 (US32) of HCMV.

Statistical Analysis

Purified peptides and protein antigens were tested in triplicate wells. For these, a paired one-tailed Student's *t*-test was performed to identify positive responses relative to the medium control wells. As ELISPOT counts follow Gaussian (normal) distribution among replicate wells, the use of parametric statistics was justified to identify positive and negative responses, respectively (21). A $p < 0.05$ was considered as the cut-off for positive responses induced by the purified peptides. The 553 individual peptides of the pp65 9-mer peptide library were tested in single wells. For these peptides, the threshold for a positive response was set at exceeding 5 SD of the mean SFU count detected in 18 replicate media control wells.

HLA-Binding Predictions

We assessed peptide-HLA I presentation by predicting peptide-HLA I binding using HLA I allele specific profile motif matrices (22–24). We considered that a given peptide binds to a specific HLA I molecule when its binding score ranks within the top 3% percentile of the binding scores computed for 1,000 random

9-mer peptides (average amino acid composition of proteins in the SwissProt database).

RESULTS

T Cells Target Multiple Antigens of HCMV

The genome of HCMV encodes multiple viral proteins, each of which could constitute a viable target antigen for T cell recognition. To this end, we tested 20 such HCMV antigens, specified in **Table 1**, for their ability to recall T cell responses in healthy human donors. Peripheral blood mononuclear cells (PBMC) obtained from six HCMV-seropositive and six HCMV-seronegative human subjects were challenged for 24 h with the specified HCMV antigens to selectively stimulate the respective antigen-specific T cell populations to secrete IFN- γ . IFN- γ production was measured in a standard ELISPOT assay format in which the cytokine is captured on the membrane around the cells that secrete it, permitting the visualization and quantification of individual IFN- γ -secreting T cells as "spot forming units" (SFU). Thus, this assay measures, at a single-cell level, the number of T cells that engaged in IFN- γ production following antigen stimulation (25). The individual HCMV antigens used

TABLE 2 | OD6+ T cell epitopes of HCMV pp65 protein recognized in A*02:01-positive healthy donors as established by brute force epitope mapping.

Donor ID	HLA Class 1			Control		Antigen-induced IFN- γ (SFU) per 1,000,000 cells														
				Medium CPI	pp65(15-mer pool)	Peptides of pp65 9-mer library														
	A	B	C			pp65(495-503) ^{1,a}	pp65(495-503) ¹	pp65(495-105) ²	pp65(97-125) ³	pp65(116-149) ⁴	pp65(141-152) ⁵	pp65(144-183) ⁶	pp65(175-211) ⁷	pp65(203-229) ⁸	pp65(221-332) ⁹	pp65(324-333) ¹⁰	pp65(325-425) ¹¹	pp65(417-426) ¹²		
284	02:01	40:08	03:04	7	1,004	175	201	0	238	0	0	0	394	0	1,145	1,327	0	0		
285	68:01	44:03	16:01																	
	02:01	35:01	06:02	11	36	14	29	0	0	0	0	0	0	0	0	0	0	0		
	03:01	45:01	16:01																	
300	02:01	39:05	02:02	11	1,681	1,114	1,074	0	0	0	0	0	0	0	0	0	0	0		
	24:02	51:01	07:02																	
331	02:01	39:05	07:02	0	533	341	458	65	80	136	446	0	0	235	0	0	0	0		
	68:03	51:01	15:09																	
350	02:01	07:24	07:01	11	2,234	40	38	0	0	0	0	0	0	0	0	0	1,496	515		
	03:01	18:01	07:02																	

¹NLPVPMATV (44.8%); ²PTGRSCPS (64.7%); ³LPLKMLNIP (64.7%); ⁴HLPLVADAVI (80.3%); ⁵VADAVIHAS (54.2%); ⁶WKEPDVYVT (54.2%); ⁷ELVCSMENT (41.2%); ⁸DOYVKVYLE (43.7%); ⁹QQIFLEVQA (41.0%); ¹⁰QIFLEVQAI (44.1%); ¹¹TPRVTTGGGA (43.3%); ¹²PRVTGGGAG (40.1%).

^aMore than 95.0% purity. SFU counts 1 \times 10⁵ PBMC are shown only for peptides that elicited recall responses with peptides being identified by amino acid positions and sequence. Positive results were defined as antigen-triggered SFU counts that exceeded 5 SD of the medium control SFU counts (the latter calculated from a total of 18 medium control wells tested per donor on six plates). While all five donors shared the HLA-A*02:01 allele, the other HLA class I alleles expressed are also listed in the table as they could also serve as restriction elements for the peptide-specific CD8+ T cells.

for stimulation were 15 amino acid (aa) long peptides that collectively spanned the respective polypeptide sequences in steps of (skipping) 11 aa, hereafter referred to as peptide pools. Each peptide was present at $\sim 1 \mu\text{g/mL}$ within the respective peptide pools, and the number of peptides contained in each pool is specified in **Table 1**.

Stimulation of all six HCMV-seronegative donors with each of the twenty HCMV peptide pools failed to elicit an increased number of IFN- γ -producing T cells relative to PBMC cultured in media alone (**Table 1**). However, each of these HCMV-seronegative donor PBMC robustly responded to a combination of cytomegalovirus (C), parainfluenza (P), and influenza (I) antigens, collectively referred to as CPI (20), which confirmed T cell functionality in the respective samples (**Table 1**). The inability to detect a recall response to the HCMV peptide pools in HCMV-seronegative donors, in the face of their CPI reactivity, establishes the exquisite specificity of the HCMV peptide pool-triggered recall responses.

Stimulation of all six HCMV-seropositive donors' PBMC, in contrast, revealed recall responses to several of these HCMV antigens (**Table 1**). T cells specific for IE-1, pp65, and UL55 were detected in all six HCMV-seropositive donors, but the magnitude of recall responses was variable between donors, and varied also within a donor, ranging from relatively low SFU counts (in the tens) to high counts (in the hundreds). As the peptide pools tested on all donors were the same, and these were tested in a single experiment, the variability of responses observed must lie in the T cell compartment itself. There was no apparent response hierarchy seen for IE-1, pp65, and UL55. The IE-2, UL28, UL32, UL36, UL82, UL94, UL103, UL153, and US3 peptide pools also elicited recall responses in at least half of these donors, and again there was no clear response hierarchy seen against these antigens. For example, Donor 64 exhibited a high frequency recall response to both the UL36 and UL55 peptide pools, with negligible responses against several other peptide pools. In contrast, the response against pp65, UL32, and US3 prevailed in Donor 99. In aggregate, IE-1, pp65, UL28, UL32, UL36, UL55, UL94, US3, and US24 peptide pools were preferentially recognized by effector T cells from these HCMV-seropositive donors. Importantly, these antigens were not consistently immunodominant, whereas IE-2, UL40, UL48-1/2, UL99, UL103, UL151, UL153, US29, and US32 were either not recognized, or recalled a low frequency of T cells in the test subjects.

Collectively, the data presented in **Table 1** demonstrate that for monitoring T cell immunity to HCMV one should not focus on a single viral antigen. Even if an antigen is immune dominant in a subset of donors, it might be subdominant or even negative in other donors who respond strongly to other antigens of the virus. Selecting one antigen over another may not only grossly underestimate the magnitude of the expressed T cell repertoire, but can also misrepresent it entirely. For example, pp65 was co-dominant with UL32 and US3 in Donor 99, whereas pp65 was co-dominant with IE-1 and UL55 in Donor 137. However, pp65 can also be a subdominant antigen. Such was the case for Donor 182, in whom the number of T cells responsive to IE-1, UL55, UL94, and US3 were each 10 times higher than those targeting pp65. Therefore, if pp65 reactivity would be used in this

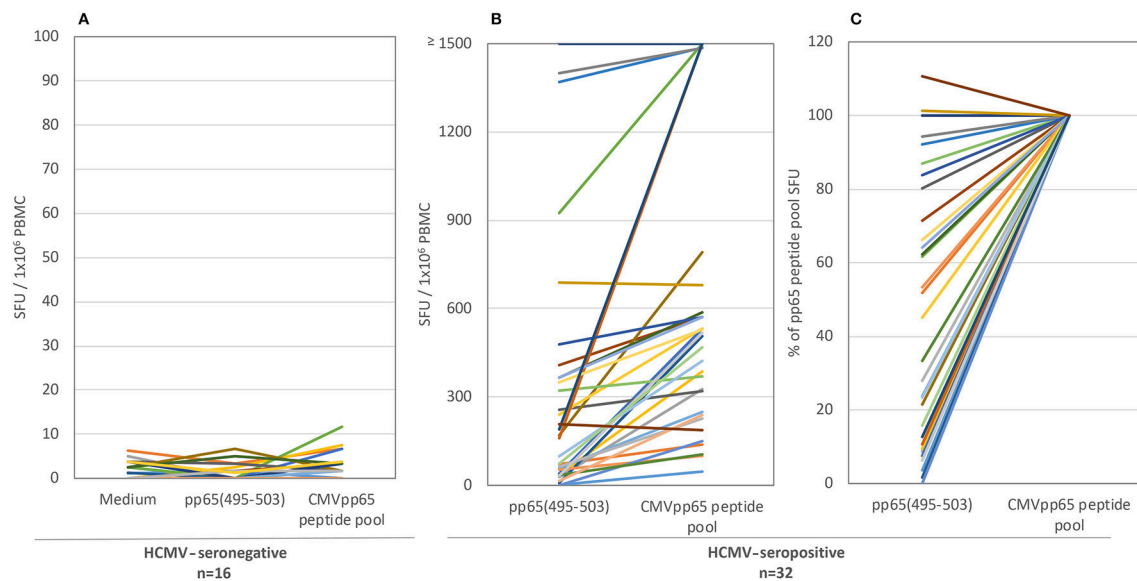


FIGURE 1 | The T cell recall response to single peptide pp65_{495–503} vs. the pp65 peptide pool in HLA-A*02:01-positive test subjects. Both antigens were tested in an IFN- γ ELISPOT assay on 16 HCMV-seronegative subjects (**A**) and 32 HCMV-seropositive donors (**B**). The results are expressed as SFU / 1×10^6 PBMC. Colored lines link each test subject's response to the two antigens. Panel (**C**) shows the normalized data for the seropositive cohort, with the SFU counts elicited by the pp65 peptide pool set as 100% for each donor, to which the magnitude of the pp65_{495–503} triggered SFU counts in the same donor was compared.

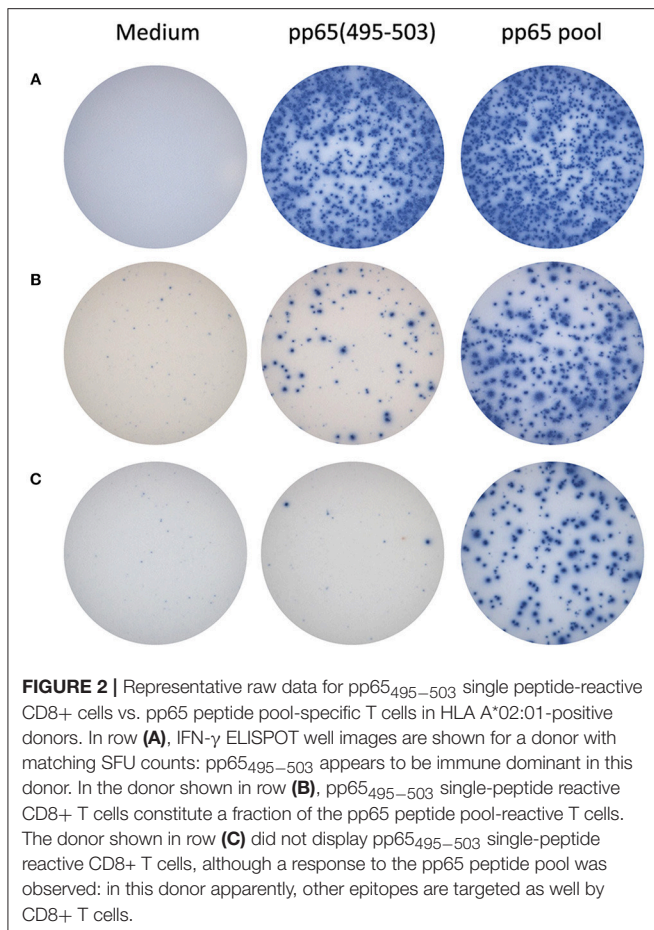
particular donor to assess the magnitude of the HCMV-specific T cell repertoire, one would detect only 1.6% of the HCMV antigen-specific T cells responsive to the epitope space covered by the 20 HCMV antigens used for testing. For HCMV, therefore, these data highlight the need to comprehensively test all antigens of the virus to reliably assess the magnitude of the anti-viral T cell response. This notion might apply to other viruses as well, since similar observations were made through testing T cell recognition of peptide libraries representing several EBV antigens (**Supplementary Table 2**).

CD8⁺ T Cell Recognition of HCMV Epitope pp65(495-503) in HLA-A*02:01-Positive Test Subjects

T cell immune monitoring is frequently performed using single peptides. Specifically, the NLVPMVATV peptide, corresponding to amino acids 495-503 of HCMV's pp65 protein (pp65_{495–503}) has been described as an immune dominant epitope of HCMV in HLA-A*02:01-positive subjects (19). Having established that restricting immune monitoring to only the pp65 antigen might be insufficient, we next sought to establish how reliably restricting immune monitoring to a single pp65 peptide would reflect on the entire pp65 epitope space. As one approach to address this question, we tested 32 HLA-A*02:01-positive, HCMV-seropositive healthy human donors for reactivity against the single pp65_{495–503} peptide and the pp65 peptide library in parallel using an IFN- γ ELISPOT assay (**Figure 1**; the raw data are shown in **Supplementary Table 3** permitting the identification of individual donors). As controls, PBMC from sixteen HLA-A*02:01-positive, HCMV-seronegative donors were

tested in the same manner. Each of these control subjects' PBMC failed to respond to either antigen (**Figure 1A**), while exhibiting a strong recall response to CPI (data not shown). The absence of a recall response to the pp65_{495–503} peptide, and the pp65 peptide pool, in HCMV-seronegative donors serves to further support the specificity of the recall response these antigens elicited in HCMV-seropositive individuals.

For the HLA-A*02:01-positive, HCMV-seropositive donor cohort ($n = 32$), the number of T cells responsive to either the pp65_{495–503} peptide or pp65 peptide pool was determined (**Figure 1B**). Since our objective was to address whether the pp65_{495–503} peptide was indeed immunodominant within the epitope space of the pp65 antigen, we normalized the number of T cells activated by the pp65 peptide pool to 100% and expressed the number of pp65_{495–503} peptide reactive T cells as a percentage of this total (**Figure 1C**). Using 50% as an arbitrary cut-off for immunodominance, 15 of these 32 (47%) donors exhibited a dominant response to the pp65_{495–503} peptide. Therefore, in approximately half of HCMV-seropositive HLA-A*02:01 subjects, reactivity to the pp65_{495–503} peptide accurately reflected the expressed T cell repertoire against the pp65 antigen. However, in 17 of the remaining HCMV-seropositive HLA-A*02:01 subjects (53%), reactivity against the pp65_{495–503} peptide constituted <50% of the pp65 peptide pool-specific T cell repertoire. In these donors, it was therefore likely that additional epitopes covered by the pp65 peptide pool were also targeted, and that these might even outnumber the pp65_{495–503}-specific T cells. Notably, in six of these HLA-A*02:01-positive, HCMV-seropositive subjects (19%), the pp65_{495–503} peptide failed to recall detectable numbers of T cells whilst responses against the pp65 peptide pool were detected. In such donors,



immune monitoring with the single pp65_{495–503} peptide, but not with the pp65 peptide pool, would yield false negative results since the expressed T cell repertoire targeted other regions of the pp65 protein. To illustrate this point further using raw data, in **Figure 2** we present representative well images of donors for which the pp65_{495–503} peptide was immune dominant, encompassing nearly the entire epitope space covered by the pp65 peptide pool (Donor 261 in Row A of **Figure 2**). In contrast, a donor in whom pp65_{495–503} specific CD8+ T cells cannot be detected in spite of this donor's reactivity to the pp65 peptide pool (Donor 183 in Row C of **Figure 2**), and a donor for which the pp65_{495–503} epitope covers only a fraction of pp65's epitope space (Donor 213 in Row B of **Figure 2**) are also presented.

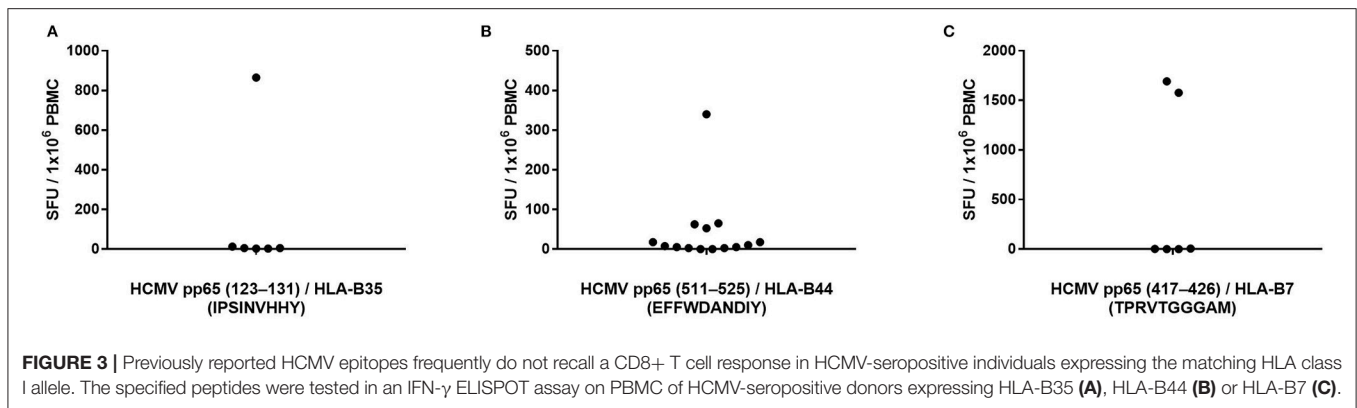
Overall, **Figures 1, 2** establish that even in those rare cases in which immune dominance of a single peptide can occur, such as pp65_{495–503} peptide recognition in HLA-A*02:01-positive subjects, the expressed antigen-specific T cell repertoire can be frequently underestimated by this single peptide, even to the point of obtaining false negative data. Furthermore, immune monitoring of T cell responsiveness to a single antigen, or peptide thereof, would fail to capture the extent to which other antigens of the same antigen system were targeted (see **Table 1**).

For the above comparisons of T cell recognition of the pp65_{495–503} peptide vs. the entire pp65 epitope space, it must be noted that the numbers of pp65_{495–503} peptide-specific CD8+ cells can be expected to be accurate, whereas the numbers obtained using the pp65 peptide pool are likely to underrepresent the entire expressed pp65 antigen-specific T cell repertoire. The pp65_{495–503} peptide is HLA-A*02:01-restricted and too short for presentation by HLA class II molecules, hence it should exclusively elicit CD8+ T cells. In contrast, the pp65 peptide pool, consisting of 15-mer peptides, recalls both CD4+ and CD8+ T cells (**Supplementary Figure 1**). Fifteen-mer peptides can trigger, but are not ideal for, CD8+ T cell activation as the peptide binding groove of HLA class I molecules is closed on both ends and therefore accommodates only peptides of 9–11 aa, and moreover is intolerant to frame shifts of the peptide binding motif (11). The 15-mer peptides therefore need to undergo further processing to generate peptides 9–10 aa in length that are suitable for binding HLA I molecules, and such processing can also destroy these peptides. Moreover, as the peptides in the pp65 pool walk the pp65 sequence in steps of 11 aa, considerable gaps in epitope coverage can be expected. The possibility that the 15-mer peptide pool insufficiently covers the entire CD8+ T cell epitope space of pp65 suggests that this pool detects only a fraction of the antigen-specific CD8+ T cell repertoire. Therefore, the difference between the single peptide data, and the full antigen-specific CD8+ T cell repertoire is potentially even larger than suggested by the data shown in **Figure 1**. To this end, for comprehensive assessment of complete CD8+ cell epitope coverage of an antigen, it would be ideal to test peptides that are 9 to 10 aa long and that cover the protein sequence in steps of single amino acids, an approach we have undertaken below.

CD8+ T Cell Recognition of HCMV Epitopes in HLA-B35-, B44-, B7-, and B18-Positive Subjects

Peptide pp65_{123–131} (IPSINVHHY) is HLA-B35-restricted and has been described as an epitope frequently targeted by CD8+ T cells from HCMV-infected donors bearing this HLA allele (20). We identified seven HLA-B35-positive, HCMV-seropositive, and pp65 peptide pool-reactive donors and tested these donors' PBMC for T cell reactivity to this peptide. As shown in **Figure 3A** (the raw data are shown in **Supplementary Table 4** permitting the identification of individual donors), only one of these seven donors' PBMC harbored pp65_{123–131}-specific T cells in high numbers. In the other six donors tested, the frequency of pp65_{123–131}-specific T cells was below the limit of detection for the assay, that is, <1 in 400,000 PBMC.

To further test whether single peptides are insufficient for accurate T cell immune monitoring, we assessed T cell reactivity against three additional pp65-derived peptides that are restricted by HLA-B alleles. The HLA-B44-restricted pp65_{511–525} (EFFWDANDIY) peptide has previously been reported as a prevalent epitope targeted by HCMV-infected individuals expressing this HLA allele (26). We identified 15 PBMC donors who were HLA-B44-positive, HCMV-seropositive, and also reacted strongly to the pp65 peptide pool. Only one



of these donors exhibited a strong recall response to the pp65_{511–525} peptide (Figure 3B; the raw data are shown in Supplementary Table 4 permitting the identification of individual donors). Three additional HLA-B44-positive donors responded weakly to the pp65_{511–525} peptide, while we failed to measure a detectable response in the remaining 11 donors (73%). The pp65_{417–426} (TPRVTGGGAM) peptide is an HLA-B7-restricted epitope that has also been implicated as a prevalently recognized HCMVpp65 epitope (27). We had accessed to PBMC from six individuals who were HLA-B7-positive, HCMV-seropositive, and pp65 peptide pool-reactive. As shown in Figure 3C (the raw data are shown in Supplementary Table 4 permitting the identification of individual donors), only two of these donors (33%) possessed pp65_{417–426}-specific T cells in high numbers, while in the remaining four donors, pp65_{417–426}-specific T cells were undetectable in spite of their HCMV-positive status. Lastly, the pp65_{378–389} (SDEEEAIVAYTL) peptide is HLA-B18-restricted and has been described as a frequently targeted epitope in HCMV-infected donors who express this HLA allele (19). We had access to three HLA-B18-positive donors who were also HCMV-seropositive and responded to the pp65 peptide pool: none of these donors displayed a recall response to the pp65_{378–389} peptide (data not shown).

Supporting the notion put forward for the pp65_{495–503} epitope in Figure 1, the data presented in Figure 3 offer further support using three additional peptides previously described as “immune dominant” epitopes: such peptides were not necessarily targeted by T cells, and thus would frequently provide false negative results, or underestimate the HCMV-specific T cell repertoire. Instead, T cells frequently recognized alternative epitopes from the pp65 antigen, as suggested by the single-peptide negative donors’ responsiveness to the pp65 peptide pool. Furthermore, pp65 is only one of many HCMV antigens recognized by the expressed T cell repertoire, and is not immunodominant in all HCMV-seropositive subjects (Table 1). Immune monitoring, therefore, that relies on single HCMV peptides is likely to provide false negative results in a considerable fraction of test subjects, and when peptide-reactive T cells are detected, their numbers might not provide an accurate reflection of the overall magnitude of the antigen-specific T cell repertoire in the test subject. We have made similar observations using the EBV and influenza

test systems (Supplementary Figures 2, 3), suggesting that this notion might hold for T cell immunity to other viruses as well.

Brute Force Epitope Mapping of HCMV pp65 Antigen

The frequent discordance between the frequency of CD8+ T cells specific for the pp65_{495–503} peptide vs. the pp65 peptide pool in HLA-A*02:01-positive, HCMV-seropositive subjects (Figure 1), along with the notion that the recall response triggered by the pp65 15-mer peptide pool encompasses both a CD8+ and CD4+ T cell component (Supplementary Figure 1) suggested that pp65_{495–503} might not be the sole immune dominant epitope in these subjects. In particular, we hypothesized that this was the case for donors in which the pp65_{495–503} peptide-induced recall response was weak relative to that triggered by the pp65 peptide pool (e.g., Donors 213 and 183 in Figure 2). To directly address this hypothesis, we set out to systematically identify all MHC class I-restricted T cell epitopes in the pp65 protein that were recognized using a brute force epitope mapping approach. Therefore, a series of 9-mer peptides was synthesized that span the entire sequence of the pp65 protein progressing in steps of a single amino acid (Supplementary Figure 4). These individual peptides were plated at 0.2 μ g/mL, one peptide sequence per well, with the 553 unique peptides spanning across six 96-well plates for each test subject. Each plate contained 3 medium only control wells, and one well allocated for a CPI positive control. In the same experiment, each donor’s PBMC (at 3×10^5 cells/well) were also tested for their recall response to the pp65 15-mer peptide pool and to pp65_{495–503} peptide (from a different synthesis, at >95% purity). Of note, the individual peptides of the 9-mer library were not subjected to further purification after their synthesis, and averaged 56% purity. The results of the IFN- γ ELISPOT assays using the individual peptides of the pp65 9-mer series, along with controls, for stimulation of four HLA-A*02:01-positive, HCMV-seropositive healthy donors (Donors 284, 300, 331 and 350) and a single HLA-A*02:01-positive HCMV-seronegative donor (Donor 285) are summarized in Table 2.

As seen in Table 2, Donor 300 responded vigorously to the 15-mer pp65 peptide pool (1,681 SFU/ 1×10^6 PBMC), and with a similar frequency to purified pp65_{495–503} peptide (1,114 SFU/ 1×10^6 PBMC). The magnitude of the

recall response to the corresponding (unpurified) pp65_{495–503} peptide from the peptide series was 1,074 SFU/ 1×10^6 PBMC, which was essentially identical to the purified pp65_{495–503} peptide. Since the recall response induced by the purified and unpurified pp65_{495–503} peptides provided highly similar SFU counts for all donors tested, this result suggested that the 9-mer pp65 peptide series was well-suited for assessing CD8⁺ cell activation despite the absence of further purification following their synthesis. Unlike other donors, Donor 300 failed to exhibit a response to any of the other peptide in the 9-mer pp65 peptide series. Therefore, the expressed T cell repertoire of Donor 300 was indeed exclusively targeting the HLA-A*02:01-restricted pp65_{495–503} peptide.

For Donor 284, the pp65_{495–503} peptide-elicited recall response was 175 SFU/ 1×10^6 PBMC for the purified peptide, and 60 SFU/300,000 PBMC for the corresponding 9-mer in the pp65 peptide series. Therefore, the response against the pp65_{495–503} peptide constituted only ~20% of the observed pp65 peptide pool response (1,004 SFU/ 1×10^6 PBMC). In agreement with this observation, Donor 284 demonstrated recall responses to several additional 9-mers in the pp65 peptide series. Specifically, the pp65_{116–125} (238 SFU/ 1×10^6 PBMC), pp65_{203–211} (394 SFU/ 1×10^6 PBMC), pp65_{324–332} (1,145 SFU/ 1×10^6 PBMC) and pp65_{325–333} (1,327 SFU/ 1×10^6 PBMC) peptides all stimulated IFN- γ secretion by this donor's PBMC. **Supplementary Figure 5** depicts raw data (well images) of the ELISPOT plate in which adjacent pp65_{324–332} and pp65_{325–333} peptides triggered a recall response in this donor. Being representative of results obtained from all plates, these well images are shown to illustrate the clarity of peptide-triggered signal over the negligible background noise in these epitope scanning assays. Since the pp65_{324–332} and pp65_{325–333} peptides are adjacent, the observed responses likely identify a single naturally processed CD8⁺ T cell epitope for this donor. Therefore, the peptide pp65_{495–503} was only one of 4 co-dominant epitopes recognized by CD8 T cells from Donor 284. Likewise, Donor 331 followed a similar pattern and recognized six distinct pp65 epitopes, of which five were identified by brute force epitope scanning and were recognized in a co-dominant fashion.

In Donor 350, the pp65_{495–503} peptide elicited recall response (40 SFU/ 1×10^6 PBMC for the purified peptide, and 38 SFU/ 1×10^6 PBMC for the corresponding nonamer from the 9-mer series) was only 2% of that triggered by the pp65 peptide pool (1,976 SFU/ 1×10^6 PBMC). For this particular donor, two adjacent peptides, pp65_{417–425} and pp65_{418–426}, that likely jointly reveal a single epitope, triggered a vigorous recall response with 1,496 and 513 SFU/ 1×10^6 PBMC, respectively. No other peptide, except for the weak recall response to pp65_{495–503} itself, triggered SFU counts above the medium background (**Table 2**). Therefore, for Donor 350 the immune dominant epitope was pp65_{417–425}/pp65_{418–426}, with the pp65_{495–503} peptide contributing only minimally to CD8⁺ T cell recognition of the pp65 antigen.

Unlike the above HCMV-seropositive donors, the HCMV-seronegative Donor 285 failed to mount a significant response to the pp65 peptide pool or pp65_{495–503} peptide. This donor also

did not yield a positive response to any of the peptides in the pp65 9-mer peptide series, supporting the specificity of responses induced by such peptides in Donors 284, 300, 331 and 350.

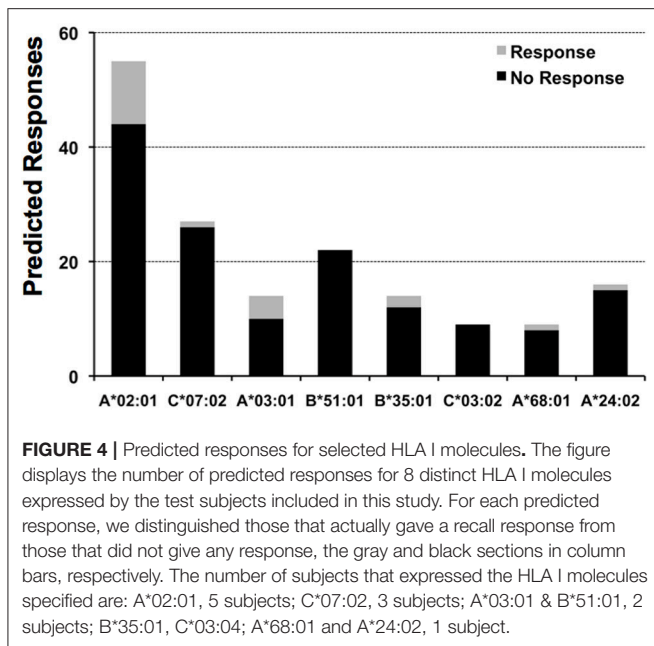
Overall, these brute force CD8⁺ epitope scanning experiments confirm the notion that immune monitoring that relies exclusively on a single “immune dominant” peptide can largely underestimate (Donors 284 and 331), or fail to detect (Donor 350) the antigen-specific CD8⁺ T cell repertoire. Only in one of four instances was the single peptide approach sufficient to accurately monitor T cell reactivity against a complex antigen (Donor 300). Therefore, to comprehensively assess the magnitude of antigen-specific CD8⁺ T cell immunity in a test subject population, it would be advisable to either test peptide pools or cover the entire epitope space using single peptides.

Predicted vs. Actually Recognized pp65 Epitopes

The epitope space evaluated for pp65 consisted of 553 peptides, 9 aa in length, spanning the entire length of the protein with overlaps of 8 aa residues. We selected 9-mer peptides since the HLA I molecules have a preference for binding peptides of that size. We tested the response of 5 subjects to each of these peptides, identifying the high frequency responses shown in **Table 2**, plus 194 weaker responses, that is, SFU counts that exceeded the mean of the 18 medium control wells by 5 SD. We assessed how many of these responses could be anticipated by determining if peptides could be predicted to bind to any of the HLA I alleles expressed by the subjects (details in Material and Methods). Of note, we would have only predicted 20 of these responses, which constitute only ~10% of experimental responses. This rate of response anticipation is likely an underestimation since we were limited by the availability of relevant profile motifs for predicting peptide-HLA I binding (detailed in Material and Methods). Thus, we could only predict peptide-HLA I binding for 8 out of 19 HLA I alleles expressed by the five test subjects detailed in **Table 2**. In any case, this rate for anticipating response is actually within the reported epitope discovery rate (10), and it is clear that most experimental responses can be predicted. Likewise, we found that positive responses represent a minority of predicted responses (**Figure 4**). In other words, we predict far more responses than we actually detected. The number of detected responses ranged from 11 out of 55 predicted for A*02:01 to none out of 22 for B*51:01. It is worth mention that the number of predicted responses is linked to both the number of peptides predicted to bind a given HLA I molecule and the number of subjects who expressed that HLA molecule. Thus, the number of predicted A*02:01 responses is larger than that for all the other HLA I molecules because all the test subjects were typed positive for A*02:01.

DISCUSSION

Using T cell reactivity against human cytomegalovirus (HCMV) Epstein-Barr virus (EBV) and influenza as prototypical examples, we present data confirming that the epitope space of these viruses comprises numerous antigens. Moreover, these data highlight the variable hierarchy in T cell recognition of these



target antigens amongst the subjects we tested (Table 1 and Supplementary Table 2). The notion that T cells target multiple viral antigens without a clear pattern of immunodominance in the human population for an individual antigen might be a generalizable finding, which is likely to apply to many other viruses as well. Therefore, immune monitoring efforts for HCMV, EBV (and possibly most other viruses), that narrows in on a single viral antigen has a high likelihood of misrepresenting the high- or low-responder status of an individual, and even of providing false negative results. Our data call attention to the need for antigen genome-wide immune monitoring efforts.

Assessment of T cell reactivity against a single, well-defined HCMV pp65 peptide (pp65_{495–503}) vs. a pool of pp65 peptides clearly demonstrated that the single peptide-induced recall response frequently underestimated the magnitude of the expressed T cell repertoire targeting the entire pp65 antigen, and in some instances failed to detect it entirely (Figures 1, 2). Likewise, assessment of T cell reactivity against additional single peptides from the pp65 antigen vs. the pp65 peptide pool further reiterated this finding (Figure 3). Similar observations were also made for EBV (Supplementary Figure 2), and for influenza (Supplementary Figure 3). Therefore, relying on single, previously-defined epitopes seems to be insufficient to reliably quantify, or even detect, T cell immunity against these three viruses, and likely against other viruses as well.

Focusing more closely on HCMV-seropositive, HLA-A*02:01-expressing test subjects which exhibited discordance between the number of T cells that responded to the “immunodominant” pp65_{495–503} peptide and the pp65 peptide library, we tried to reconcile this discrepancy by testing whether such donors possessed CD8+ T cells that recognized alternative pp65 epitopes. We sought to test this hypothesis through “brute force” epitope scanning of the entire pp65 antigen, using individual

9-mer peptides that span the entire pp65 aa sequence. Indeed, each of these test subjects responded to additional pp65-derived 9-mer peptides and no single 9-mer peptide was universally recognized by each of these donors (Table 2). Taken together, these data further highlight the need to cover epitopes of an antigen comprehensively through usage of peptide libraries, rather than relying on a single or few previously defined “immune dominant” peptides.

One possible solution for performing comprehensive CD8+ T cell immune monitoring is to tailor peptide libraries to specific HLA class I alleles. Existing algorithms can be used to predict likely peptides that encompass the potential epitope space recognized by CD8+ T cells based on HLA binding criteria (28). However, considering each HLA class I allele expressed by individual test subjects, and a complex antigen such as pp65, this is a considerable scope. There will be a multitude of predicted peptides that satisfy the imposed HLA binding criteria, and these peptides will in large be different for each donor. As a much simpler alternative to such designed peptide libraries, one can take the “agnostic” brute force approach in which a peptide library is constructed that walks the entire amino acid sequence in steps of single amino acids. Using this latter method, one can conceivably cover the entire antigenic space of a protein, in any human, without the need for customization. This brute force approach can actually identify peptide responses where prediction fails, as only a minority of the responses can actually be predicted, and moreover, only a minority of predicted responses can actually be detected.

For a “brute force” approach, 9–11 aa peptides have been successfully used to detect epitope-specific CD8+ T cells (13, 14, 29). In these studies, similar to the results presented in this communication, frequently just one or two adjacent peptides will elicit a CD8+ T cell recall response (Table 2 and Supplementary Figure 5) (13, 14, 29). This natural law confirmed here, that is dictated by the closed peptide-binding groove of MHC class I molecules, calls into question the use of peptide libraries consisting of peptides longer than 11 aa, and that walk the protein sequence in steps greater than a single aa. Thus, only comprehensive coverage of all potential epitopes can reveal the exact dimensions of the expressed CD8+ T cell repertoire.

Therefore, in theory, testing of peptide libraries consisting of 9–11 aa length with single amino acid overlaps would be the ideal strategy for “agnostic” CD8+ T cell immune monitoring. However, *in praxi*, until recently such an approach would have been considered impractical and prohibitive. This is due to (a) the number of peptides needed for such an approach, (b) the number of primary lymphoid cells (e.g., PBMC) required, (c) the labor involved in such testing, (d) the scope of data analysis and (e) interpretation of the flood of data obtained. As we illustrated in this work through systematic testing of five donors’ PBMC responding to 553 single peptides, all tested in one experiment, these limitations no longer apply.

The primary notion set forth in this communication is that genome-wide comprehensive testing of 9–10 amino acid long peptides is both necessary and feasible for CD8+ T cell immune monitoring; restricting it to a few known “immune dominant” peptides is likely to underestimate or even entirely miss the

antigen-specific CD8+ cell repertoire. Immune monitoring of CD8+ cells critically depends on the precise use of peptides, while immune monitoring of CD4+ cells can be done without involving peptides, using the entire protein. Our data also support the notion that large scale peptide scans have become feasible for systematic epitope discovery. Major steps have already been undertaken in this direction for CD4+ T cells (30, 31). Because HLA Class II molecules' peptide-binding grooves are open on both ends, they can both accommodate longer peptides, and are tolerant to frame shifts of the peptides' HLA anchor residues. For this reason, one can expect CD4+ cell determinant mapping to be largely comprehensive when the classic approach is taken using longer peptides (e.g., 15–25-mers) that cover the antigen sequence in larger steps (e.g., 5–15 amino acid increments). This classic approach for CD4+ cells would likely miss many or even most CD8+ cell epitopes, however (see **Table 1**, and **Supplementary Figure 5**). Systematically covering all possible non-amer peptides for mapping CD8+ T cell determinants multiplies the number of peptides to be tested, and thus the scope of testing. Our data suggest that, even for CD8+ T cells, comprehensive large-scale genome wide epitope discovery is approaching feasibility.

Recent advances in high-throughput peptide synthesis technologies have enabled manufacturers to offer extensive, even full proteome-spanning peptide libraries at sufficient quality and low cost, including the customization of such libraries. Unpurified peptides are suitable for screening purposes because the HLA alleles and TCR present in the PBMC cultures can be expected to select the “right” peptide synthesis variants that primed the T cell responses *in vivo*. Confirming this notion, we found (see **Table 2**) that in all five donors tested the purified pp65_{495–503} peptide recalled essentially identical numbers of CD8+ T cells as the corresponding unpurified peptide from the 9-mer pp65 peptide series. However, as crude peptides are variable in both their purity level and yield, in general, it is advisable to validate individual epitopes identified in such screening experiments using more stringently purified peptides. Thus, the availability and manufacture of extensive peptide libraries no longer precludes brute force T cell epitope mapping.

The greatest obstacle, and rate-limiting factor, for brute force epitope mapping is the number of PBMC that can be obtained from a single test subject. For testing 553 peptides at 3×10^5 PBMC/well, plus 23 control wells, we required 1.73×10^8 PBMC from each test subject. Up to 5×10^8 PBMC can be obtained by classic venipuncture. As an alternative, which we relied upon, one can readily obtain 2×10^{10} , or more, PBMC from a human subject in a single leukapheresis draw while depleting that individual of only 1% of his/her white blood cells. With 2×10^{10} PBMC, using ELISPOT, one has sufficient cells to test up to 6×10^5 individual peptides!

When ELISPOT assays are performed in 96-well plates, between 0.1 and 1×10^6 PBMC are plated per well. In this cell density range, the SFU counts (e.g., the number of antigen-specific T cells detected) is strictly linear to the number of APC plated (32). In the 96-well format, 1×10^5 PBMC per well is the lowest cell input that yields reliable data. Lower numbers of PBMC in the 96-well assay no longer form a

confluent cell layer at the bottom of the well, and thus predictable T cell-APC interactions become disrupted. Recently, 384-well ELISPOT plates have become available. The membrane size on the bottom of a 384-well ELISPOT plate is one-third (not the expected one-fourth) that of the 96-well plate. The 384-well format permits miniaturization of ELISPOT assays to precisely one-third compared to a 96-well plate, whereby plating precisely one-third the numbers of PBMC yields precisely one-third of the SFU counts (32). Thus, epitope mapping experiments could theoretically be performed using 3×10^4 PBMC per well (per peptide). Based on these basic parameters, using 5×10^7 cells (corresponding to 50 mL of blood), one can test 500 individual peptides in a 96-well plate format. Using 2×10^{10} PBMC, acquired through leukapheresis, 2×10^5 individual peptides can be tested in the 96-well format with 1×10^5 PBMC per well. The number of peptides that can be tested individually in a 384-well format follows by multiplying the above numbers by 3, that is, using the 384-well format, as many as 6×10^5 individual peptides can be tested against PBMC from a donor following a single leukapheresis. Thus, similar to the number of peptides, the number of PBMC required for large scale brute force epitope mapping is not an insurmountable limitation.

Testing of large numbers of peptides in a T cell assay requires well-developed logistics. In our case, the peptides were delivered by the manufacturer as powder in the desired well format spread across six individual 96-well plates. The peptides were then dissolved “en block” using a 96-well multichannel pipettor. In this way, master peptide plates were conveniently created from which the peptides were then transferred, again “en block,” into the actual assay plates. Using such a strategy, both the dilution and plating of peptides was fast and fail-proof since the peptide layouts of the master plates were preserved. The PBMC were also added using the 96-well multichannel pipettor to conclude the most labor-intensive components of the assay: the cell culture setup. After a 24 h *in vitro* incubation, during which peptide-specific T cells became activated and secreted IFN- γ following interaction with APC, the 96-well plates were washed to remove cellular material, and detection reagents were added to begin development of the assay, all done by 96-well pipetting.

Using the approach outlined above, the setup of the experiment summarized in **Table 2**, in which 553 individual peptides plus controls were tested individually on five test subjects, took three investigators ~3 h of work to accomplish. Using this “en block” pipetting approach, theoretically, even the extreme of testing 6×10^5 peptides in 384-well format, requiring 1,563 plates, could reasonably be achieved with the assistance of pipetting robots.

As illustrated by the data presented herein, epitope recognition by T cells typically involves considerable inter-individual variability, even within HLA-allele matched donor cohorts. Therefore, systematic study of the recognized epitope space of a complex antigen system, such as a virus, will likely require testing of sizable cohorts. Because of the ease of peptide and PBMC plating when the above logistics are followed, PBMC of several donors can be tested in a single experiment. Therefore, performing the T cell ELISPOT assay itself is also

not a rate limiting factor for high-throughput, brute force epitope mapping.

The logistics of reading and analyzing ELISPOT data generated in high-throughput epitope mapping experiments involves linking peptide and test subject identities with the experimental data itself. This can be accomplished using the SpotMap™ software, which permits assignment of individual peptides and PBMC donors to specific wells of 96- or 384- well plates, including usage of unique, barcode-based plate identifiers. Using this approach, the specific plate layout can be carried through from peptide synthesis, peptide transfer, plate reading and data analysis. If done in this manner, acquisition and analysis of the raw data, including assignment of positive wells and identification of the specific peptide responsible for cytokine production, is streamlined and in large fully automatic. With the assistance of a plate stacker for the thirty 96-well plates thirty 96-well plates required to generated the data presented in **Table 2**, the scanning, counting and analysis of the 14,400 wells (corresponding to 14,400 test conditions/data point), took 40 min of fully automatic reader time. With this analysis process progressing at a rate of 1 min per 96-well plate, even the “herculean” task of analyzing the results of testing 600,000 peptides in 1,563 384-well assay plates could be accomplished within 26 h of automated reader time.

CONCLUSION

The data presented in this report not only affirm the need for experimental epitope verification, but also serve to make the point that such experiments have become technically feasible. Testing of 553 individual peptides, using five PBMC donors in a single experiment, as reported herein, represents a milestone accomplishment in that direction. We would suggest that, using the technology presented here, it should be possible to epitope map entire genomes, e.g., of viruses. Thereby, for the first time, it will become feasible to study expressed T cell repertoires recognizing entire antigenic systems. In the coming years, we anticipate that such high-resolution studies of individuals, and cohorts, epitope space should become a reality, and will fertilize the field of “epitomics.” Similar to other recently-developed

“omics” platforms, the abundance and speed with which new information will be acquired, while overwhelming to the human mind, is destined to shed unprecedented insight into T cell epitope recognition in the context of both health and disease, and eventually will permit the precise assessment of the antigen-specific T cell repertoire as required for accurate immune monitoring of CD8+ T cell immunity.

ETHICS STATEMENT

The white blood cells used in this study were collected under an IRB of Hemacare, Van Nuys, California. They were sold anonymously to CTL, that is CTL, and the authors have no knowledge of the identity of these PBMC donors. Under such conditions, the third party (i.e., CTL and the authors) does not need an IRB to use such cells for research purposes.

AUTHOR CONTRIBUTIONS

Experiments were designed by PL, AK, and PR. Experimental data were generated by MS, TZ, DR, GK, and AL, and peptide-binding analysis was performed by PR. PL, GK, and PR prepared the manuscript.

FUNDING

This study was funded by the R&D budget of Cellular Technology Limited.

ACKNOWLEDGMENTS

We thank Ruliang Li of Cellular Technology Limited for expert technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.00655/full#supplementary-material>

REFERENCES

1. Bridgeman JS, Sewell AK, Miles JJ, Price DA, Cole DK. Structural and biophysical determinants of alphabeta T-cell antigen recognition. *Immunology*. (2012) 135:9–18. doi: 10.1111/j.1365-2567.2011.03515.x
2. Maenaka K, Jones EY. MHC superfamily structure and the immune system. *Curr Opin Struct Biol*. (1999) 9:745–53. doi: 10.1016/S0959-440X(99)00039-1
3. Germain RN. MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation. *Cell*. (1994) 76:287–99. doi: 10.1016/0092-8674(94)90336-0
4. Klein J, Sato A. The HLA system. First of two parts. *New Engl J Med*. (2000) 343:702–9. doi: 10.1056/NEJM200009073431006
5. Reche PA, Reinherz EL. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. *J Mol Biol*. (2003) 331:623–41. doi: 10.1016/S0022-2836(03)00750-2
6. Nagy ZA, Lehmann PV, Falcioni F, Muller S, Adorini L. Why peptides? Their possible role in the evolution of MHC-restricted T-cell recognition. *Immunol Today*. (1989) 10:132–8. doi: 10.1016/0167-5699(89)90247-8
7. Welsh RM, Che JW, Brehm MA, Selin LK. Heterologous immunity between viruses. *Immunol Rev*. (2010) 235:244–66. doi: 10.1111/j.0105-2896.2010.00897.x
8. Yewdell JW, Bennink J. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Ann Rev Immunol*. (1999) 17:51–88. doi: 10.1146/annurev.immunol.17.1.51
9. Moldovan I, Targoni O, Zhang W, Sundararaman S, Lehmann PV. How frequently are predicted peptides actually recognized by CD8 cells? *Cancer Immunol Immunother*. (2016) 65:847–55. doi: 10.1007/s00262-016-1840-7
10. Frahm N, Yusim K, Suscovich TJ, Adams S, Sidney J, Hrabar P, et al. Extensive HLA class I allele promiscuity among viral CTL epitopes. *Eur J Immunol*. (2007) 37:2419–33. doi: 10.1002/eji.200737365

11. Rammensee HG, Falk K, Rotzschke O. Peptides naturally presented by MHC class I molecules. *Annu Rev Immunol.* (1993) 11:213–44. doi: 10.1146/annurev.iy.11.040193.001241
12. Cao K, Hollenbach J, Shi X, Shi W, Chopek M, Fernandez-Vina MA. 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. *Hum Immunol.* (2001) 62:1009–30. doi: 10.1016/S0198-8859(01)00298-1
13. Anthony DD, Valdez H, Post AB, Carlson NL, Heeger PS, Lehmann PV. Comprehensive determinant mapping of the hepatitis C-specific CD8 cell repertoire reveals unpredicted immune hierarchy. *Clin Immunol.* (2002) 103(3 Pt 1):264–76. doi: 10.1006/clim.2001.5193
14. Heeger PS, Valujskikh A, Lehmann PV. Comprehensive assessment of determinant specificity, frequency, and cytokine signature of the primed CD8 cell repertoire induced by a minor transplantation antigen. *J Immunol.* (2000) 165:1278–84. doi: 10.4049/jimmunol.165.3.1278
15. Zhang W, Lehmann PV. Objective, user-independent ELISPOT data analysis based on scientifically validated principles. *Methods Mol Biol.* (2012) 792:155–71. doi: 10.1007/978-1-61779-325-7_13
16. Franzoni G, Kurkure NV, Essler SE, Pedrera M, Everett HE, Bodman-Smith KB, et al. Proteome-wide screening reveals immunodominance in the CD8 T cell response against classical swine fever virus with antigen-specificity dependent on MHC class I haplotype expression. *PLoS ONE.* (2013) 8:e84246. doi: 10.1371/journal.pone.0084246
17. Lewinsohn DM, Swarbrick GM, Cansler ME, Null MD, Rajaraman V, Frieder MM, et al. Human *Mycobacterium tuberculosis* CD8 T cell antigens/epitopes identified by a proteomic peptide library. *PLoS ONE.* (2013) 8:e67016. doi: 10.1371/journal.pone.0067016
18. Ramachandran H, Laux J, Moldovan I, Caspell R, Lehmann PV, Subbramanian RA. Optimal thawing of cryopreserved peripheral blood mononuclear cells for use in high-throughput human immune monitoring studies. *Cells.* (2012) 1:313–24. doi: 10.3390/cells1030313
19. Wills MR, Carmichael AJ, Mynard K, Jin X, Weekes MP, Plachter B, et al. The human cytotoxic T-lymphocyte (CTL) response to cytomegalovirus is dominated by structural protein pp65: frequency, specificity, and T-cell receptor usage of pp65-specific CTL. *J Virol.* (1996) 70:7569–79.
20. Schiller A, Zhang T, Li R, Duechting A, Sundararaman S, Przybyla A, et al. A positive control for detection of functional CD4 T cells in PBMC: the CPI pool. *Cells.* (2017) 6:E47. doi: 10.3390/cells6040047
21. Karulin AY, Caspell R, Dittrich M, Lehmann PV. Normal distribution of CD8+ T-cell-derived ELISPOT counts within replicates justifies the reliance on parametric statistics for identifying positive responses. *Cells.* (2015) 4:96–111. doi: 10.3390/cells4010096
22. Reche PA, Glutting JP, Reinherz EL. Prediction of MHC class I binding peptides using profile motifs. *Hum Immunol.* (2002) 63:701–9. doi: 10.1016/S0198-8859(02)00432-9
23. Reche PA, Glutting JP, Zhang H, Reinherz EL. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. *Immunogenetics.* (2004) 56:405–19. doi: 10.1007/s00251-004-0709-7
24. Reche PA, Reinherz EL. Prediction of peptide-MHC binding using profiles. *Methods Mol Biol.* (2007) 409:185–200. doi: 10.1007/978-1-60327-118-9_13
25. Hesse MD, Karulin AY, Boehm BO, Lehmann PV, Tary-Lehmann M. A T cell clone's avidity is a function of its activation state. *J Immunol.* (2001) 167:1353–61. doi: 10.4049/jimmunol.167.3.1353
26. Longmate J, York J, La Rosa C, Krishnan R, Zhang M, Senitzer D, et al. Population coverage by HLA class-I restricted cytotoxic T-lymphocyte epitopes. *Immunogenetics.* (2001) 52:165–73. doi: 10.1007/s002510000271
27. Weekes MP, Carmichael AJ, Wills MR, Mynard K, Sissons JG. Human CD28-CD8+ T cells contain greatly expanded functional virus-specific memory CTL clones. *J Immunol.* (1999) 162:7569–77. doi: 10.1046/j.1365-2567.1999.00901.x
28. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics.* (2009) 61:1–13. doi: 10.1007/s00251-008-0341-z
29. Targoni OS, Lehmann PV. Endogenous myelin basic protein inactivates the high avidity T cell repertoire. *J Exp Med.* (1998) 187:2055–63. doi: 10.1084/jem.187.12.2055
30. Gideon HP, Wilkinson KA, Rustad TR, Oni T, Guio H, Sherman DR, et al. Bioinformatic and empirical analysis of novel hypoxia-inducible targets of the human antituberculosis T cell response. *J Immunol.* (2012) 189:5867–76. doi: 10.4049/jimmunol.1202281
31. Tian Y, da Silva Antunes R, Sidney J, Lindestam Arlehamn CS, Grifoni A, Dhanda SK, et al. A review on T cell epitopes identified using prediction and cell-mediated immune models for *Mycobacterium tuberculosis* and *Bordetella pertussis*. *Front Immunol.* (2018) 9:2778. doi: 10.3389/fimmu.2018.02778
32. Hanson J, Sundararaman S, Caspell R, Karacsony E, Karulin AY, Lehmann PV. ELISPOT assays in 384-well format: up to 30 data points with one million cells. *Cells.* (2015) 4:71–83. doi: 10.3390/cells4010071

Conflict of Interest Statement: PV, MS, TZ, DR, GK, AK, and AL are employees of Cellular Technology Limited (CTL), a company that specializes in ELISPOT testing, producing high-throughput-suitable readers, test kits, and GLP-compliant contract research.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Lehmann, Suwansaard, Zhang, Roen, Kirchenbaum, Karulin, Lehmann and Reche. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Application of Modeling Approaches to Explore Vaccine Adjuvant Mode-of-Action

Paul R. Buckley^{1,2}, Kieran Alden², Margherita Coccia³, Aurélie Chalon³, Catherine Collignon³, Stéphane T. Temmerman³, Arnaud M. Didierlaurent³, Robbert van der Most³, Jon Timmis^{2,4}, Claus A. Andersen^{3*} and Mark C. Coles^{1*}

¹ Kennedy Institute of Rheumatology, University of Oxford, Oxford, United Kingdom, ² Department of Electronic Engineering, University of York, York, United Kingdom, ³ GSK, Rixensart, Belgium, ⁴ Faculty of Technology, University of Sunderland, Sunderland, United Kingdom

OPEN ACCESS

Edited by:

Rino Rappuoli,
GlaxoSmithKline, Italy

Reviewed by:

Thorsten Demberg,
Marker Therapeutics, United States
Michael Schotsaert,
Icahn School of Medicine at Mount
Sinai, United States

*Correspondence:

Claus A. Andersen
claus.a.andersen@gsk.com
Mark C. Coles
mark.coles@kennedy.ox.ac.uk

Specialty section:

This article was submitted to
Vaccines and Molecular Therapeutics,
a section of the journal
Frontiers in Immunology

Received: 19 February 2019

Accepted: 27 August 2019

Published: 12 September 2019

Citation:

Buckley PR, Alden K, Coccia M,
Chalon A, Collignon C,
Temmerman ST, Didierlaurent AM,
van der Most R, Timmis J,
Andersen CA and Coles MC (2019)
Application of Modeling Approaches
to Explore Vaccine Adjuvant
Mode-of-Action.
Front. Immunol. 10:2150.
doi: 10.3389/fimmu.2019.02150

Novel adjuvant technologies have a key role in the development of next-generation vaccines, due to their capacity to modulate the duration, strength and quality of the immune response. The AS01 adjuvant is used in the malaria vaccine RTS,S/AS01 and in the licensed herpes-zoster vaccine (Shingrix) where the vaccine has proven its ability to generate protective responses with both robust humoral and T-cell responses. For many years, animal models have provided insights into adjuvant mode-of-action (MoA), generally through investigating individual genes or proteins. Furthermore, modeling and simulation techniques can be utilized to integrate a variety of different data types; ranging from serum biomarkers to large scale “omics” datasets. In this perspective we present a framework to create a holistic integration of pre-clinical datasets and immunological literature in order to develop an evidence-based hypothesis of AS01 adjuvant MoA, creating a unified view of multiple experiments. Furthermore, we highlight how holistic systems-knowledge can serve as a basis for the construction of models and simulations supporting exploration of key questions surrounding adjuvant MoA. Using the Systems-Biology-Graphical-Notation, a tool for graphical representation of biological processes, we have captured high-level cellular behaviors and interactions, and cytokine dynamics during the early immune response, which are substantiated by a series of diagrams detailing cellular dynamics. Through explicitly describing AS01 MoA we have built a consensus of understanding across multiple experiments, and so we present a framework to integrate modeling approaches into exploring adjuvant MoA, in order to guide experimental design, interpret results and inform rational design of vaccines.

Keywords: vaccines, adjuvants, mathematical modeling, computational biology, systems biology, mechanistic modeling, AS01

INTRODUCTION

Adjuvants are immunostimulants that shape and enhance the immune response to antigens through mimicking key aspects of innate pathogen recognition, leading to robust long-term memory recall responses (1, 2). Many modern vaccine adjuvants activate pattern-recognition-receptors (PRRs) expressed on innate immune cells, including toll-like-receptors (TLRs) and NOD-like-receptors (NLRs) (3–5) although there are a breadth

of potential innate activation mechanisms (1, 3–6). This capacity to enhance responses not only increases efficacy but can reduce the required quantity of antigen in vaccine formulations, enhancing supply in the case of pandemic infections (3). Thus, understanding how adjuvants modulate the immune response is key to providing a mechanism-based approach to rationally tailor vaccines. While non-adjuvanted vaccines are *usually* capable of inducing sufficient antibody responses, it is widely characterized that adjuvants are capable of enhancing and altering the quality of humoral responses (7). Furthermore, for some diseases, including malaria, HIV, and TB, antibody responses alone are not considered to be sufficient to eliminate the pathogen (8). Thus, adjuvants which can generate both robust CD4+ T cell and strong neutralizing antibody responses are required (9).

AS01 is a liposome-based vaccine-adjuvant containing two immunostimulants: monophosphoryl lipid A (MPL) and QS-21 (Antigenics LLC, a wholly owned subsidiary of Aenus Inc., a Delaware, USA corporation). MPL is a TLR4 agonist, and QS-21 is a saponin, derived from the *Quillaja saponaria* soap bark tree. This formulation has been shown to enhance both antibody and T helper 1 (Th1) responses to antigens (8, 10). It is currently employed in two approved vaccines against malaria and herpes-zoster virus (9). AS01-adjuvanted vaccines have shown high efficacy results in herpes-zoster phase III trials, where two doses result in >90% efficacy against herpes-zoster, regardless of age, including in >70 year old patients (11, 12), showcasing an ability to overcome the age-related defect associated with vaccination. Furthermore, in a phase 2b clinical trial, AS01-adjuvanted vaccines provided 54% protection against active TB (13). TLR agonists like MPL are often utilized in modern adjuvants, as they modulate the type and duration of the immune response (14–17). However, adjuvants can be greatly improved by the inclusion of additional immunostimulants, as observed with AS01, where, QS-21 is found to synergise with MPL. While MPL's mode-of-action (MoA) is widely characterized, QS21's MoA is not well-understood, although it has been shown to co-localize with subcapsular-sinus macrophages (SCS-M) leading to inflammasome activation in a caspase-1 dependent manner (18–21). Caspase-1 activation can trigger pyroptosis, activation of damage-associated-molecular patterns (22) and cleave pro-IL-1 β and pro-IL-18 into bioactive pro-inflammatory IL-1 β and IL-18 (23, 24). IL-1 β is pleiotropic in function, rather IL-18 has a specific role in innate IFN γ production (25). AS01-adjuvanted vaccines induce an inflammatory response associated with transient production of innate IFN γ in the draining lymph nodes of mice, peaking at approximately 6 h post-injection (PI), and subsiding to baseline levels by 48 h PI (8). It has been shown that this early IFN γ production is required to promote a functional CD4+ T-cell response in mouse models (10). Furthermore, in humans, serum IFN γ is associated with clinical protection (22). However, some key questions remain regarding the MoA of AS01, such as the role of key early events in the adaptive response.

Genetically modified animal models have provided vital insight into the mechanistic processes underpinning vaccine efficacy. Over the past two decades, these models have been utilized to determine AS01 MoA. These pre-clinical models have

permitted investigation of individual genes and proteins, in a reductionist manner (26), however not all mechanistic questions can be addressed in this way. Systems biology methodologies including machine learning, statistical, mathematical, and agent-based models can provide a holistic perspective on MoAs through data and knowledge integration (27–30). This can permit exploration of the relationships between different components in the biological system through simulation, where systems are not viewed purely as a sum of parts, but where additional phenomena can emerge as a result of integration. These methods can capture the complexity of the biological system allowing exploration of individual or population dynamics, the role of localized microenvironments, vaccination dose and time (31), and can be used to guide and optimize experimental design (27, 30, 32–36). This permits exploring dose modulation, prioritizing research avenues and determining experimental endpoints that maximize the value of individual animal experiments. Different systems-based approaches are increasingly being applied to biomedical research problems; permitting development of novel mechanistic hypotheses, spatio-temporal analysis of function of cytokines, chemokines, growth-factors, and cell-cell interactions that currently cannot be achieved *in vivo* (34–38).

Yet if systems-based modeling approaches are to add value to our understanding of the biological system, it is critical that the relationship between the biological understanding and how this knowledge is captured *in-silico* is understood. In the realms of immunology, our previous work has shown the adoption of a principled approach to the development of such tools, focusing on developing confidence that the model is fit for its purpose as providing a platform for exploring and contributing to our understanding of real-world biological systems (30, 35, 39–43).

These concepts, however, have rarely been applied to exploring adjuvant MoAs, which are highly complex systems, spanning multiple organs, and levels of biological hierarchy. Thus, we present a framework in which we follow a principled modeling process (44) and collate knowledge surrounding AS01 MoA, which will then be used to construct simulations to explore key mechanisms of interest. We have captured our current consensus of AS01 MoA (see **Figure 1**) through interdisciplinary teamwork detailing the functionality of components underpinning how AS01 works, with focus on how the production of innate-IFN γ drives an adaptive response. This work, where a system of interest is identified, modeled, and scientific questions are elucidated, collectively comprises the “Discovery Phase” of the modeling process (44). The result is a “Domain Model” which is a model (i.e., an abstraction) of the key biological detail (44), which serves as a biological basis for simulation construction (**Figure 1E**). Domain models describe only the relevant biology, and do not describe concepts related to simulation construction, or how computer code is developed (44). Decisions on which modeling methods are utilized are taken subsequently during development of a separate model, named a “Platform Model,” where mathematical and computational concepts are introduced, detailing how the Domain Model is to be implemented as a simulation (44). In this perspective, we focus on the presentation of a framework for creating a non-executable Domain Model that captures and

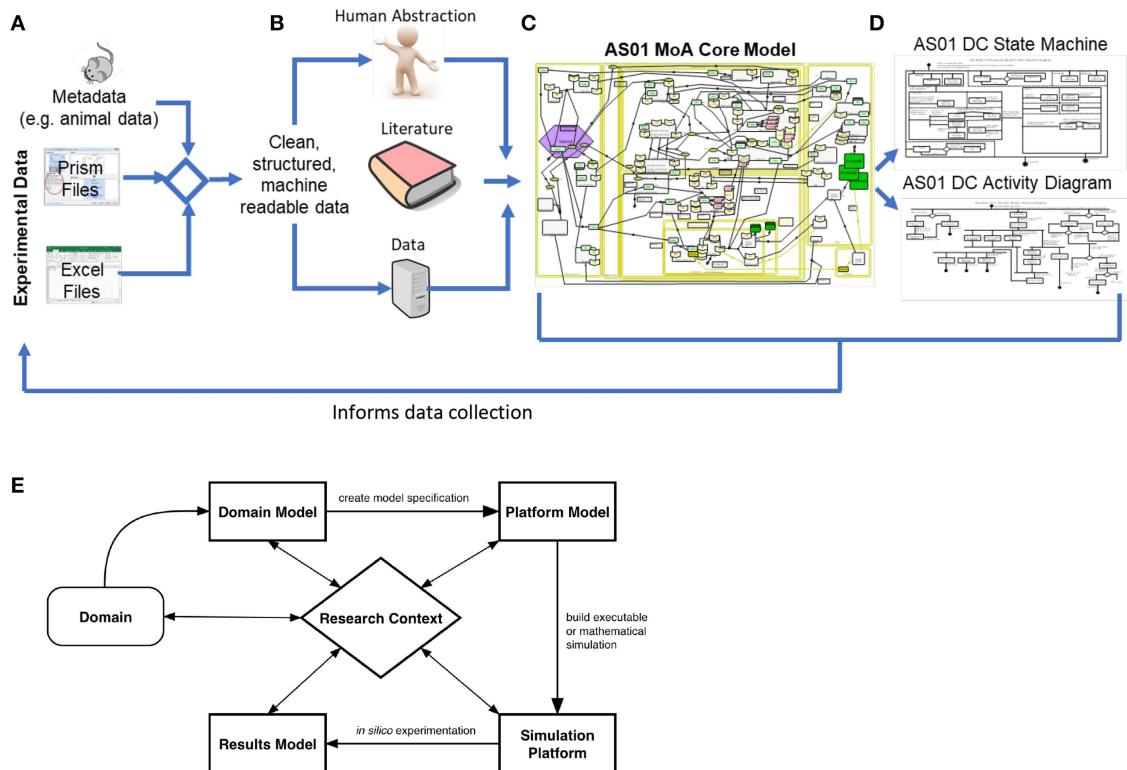


FIGURE 1 | Development of the Domain Model: in this context, development of the domain model began with **(A)** the collection of experimental data from Graphpad Prism and Excel files, and the corresponding experiment metadata. Following this, an automated pipeline was developed to clean, structure, and integrate the experimental results and the corresponding metadata, resulting in machine-readable data. Next, human abstraction, the AS01 data and immunological literature were appropriately integrated **(B)**. The AS01 data informed the development of the model 2-fold, firstly by constituting the available evidence behind the captured processes, but also by informing the rationale underpinning any human abstraction required to inform gaps in understanding. The core MoA model **(C)** was utilized to scope the requirements of the lower-level models describing component function, and so from the core MoA model, state-machine and activity diagrams were developed **(D)**. Furthermore, Domain Model development can inform data collection by permitting the understanding of what data is required to ask the scientific question. **(E)** An overview of the CoSMoS process. The system of interest, in this context a biological system, is known as the domain. From this, the current scientific understanding of the system is modeled with respect to the research context, and encapsulated in the “Domain Model”. From this, and often after a process of refinement driven by the scientific question, a software blueprint called a Platform Model is developed. At this stage, further abstractions may be necessitated, based on available computational resources or model simplifications, which are decisions that should be documented as part of the Platform Model. Following this rigorous process builds confidence in the model and simulation, improving the likelihood that the simulation is a “fit for purpose” representation of the biological system (43–45). The simulation platform is an executable piece of software written in computer code, that implements the Platform Model, while the results model encapsulates the understanding that is generated from the simulation.

brings together understanding of AS01 biology, and present specific exemplars describing key component functions. To embrace simplicity, focus was placed on capturing essential components and entities where ample evidence of involvement in AS01 MoA is available. We believe that the application of this framework will complement work to explore AS01 MoA, and that these concepts should be utilized more generally to further understand adjuvant MoAs and thus enhance vaccine efficacy.

CAPTURING THE HIGH-LEVEL CONSENSUS OF AS01 MOA IN A CELLDISIGNER MODEL

Based on murine experimental data and a consensus understanding of AS01 knowledge we have developed a

hypothesis of AS01’s MoA. This was initiated with a focus on key questions in AS01 biology: “*In preclinical models, after intramuscular injection of a vaccine adjuvanted with AS01, how does AS01 initiate an immune response? What are the key interactions that give rise to the generation of an antigen-specific CD4+ T helper cell responses and subsequent antibody production? How does IFN γ regulate these different processes?*” To address these questions, we identified and captured key cells and processes in the model. The highly visual characteristics of the Systems Biology Graphical Notation (SBGN) (46) applied in CellDesigner (47) (www.celldesigner.org) allows for improved communications which is crucial when working in cross-disciplinary teams; permitting transparency, knowledge retention, and future reusability through linking AS01 data and wider immunological literature via diagrams of biological processes. The development process, which is described below,

evolved a joint understanding between research teams, resulting in one shared model (**Figure S1**). This process permitted the confirmation of knowledge gaps, and the resulting consensus of AS01 MoA will inform the remaining development of the Domain Model. The following four step process (**Figure 1A–C**) was used to capture AS01 MoA in a process diagram: (1) Development of a biological “cartoon” incorporating current knowledge of how the adjuvant generates an adaptive immune response in a specific host; (2) The “cartoon” was used to generate a formal CellDesigner model; (3) An iterative process of CellDesigner model development was followed, to incorporate key team ideas, capturing, refining, and extending key aspects of the biology; (4) The final model was scrutinized, based on team discussions, resulting in a collective understanding agreed by all parties, to generate a single combined model of the adjuvant’s MoA.

A DESCRIPTION OF THE AS01 CELLDISIGNER MODEL

While the CellDesigner model does describe some detail (these differences are depicted in yellow) of the response after a booster dose of an AS01-adjuvanted vaccine, its main focus is the primary response. Thus, the following section describes the generation of a primary murine response to an antigen adjuvanted with AS01. As observed *in vivo*, in the CellDesigner model, after intramuscular injection of an AS01-adjuvanted-vaccine (**Figure S1A**) the adjuvant components both activate local cells and drain into the dLN (**Figures S1C–E**) initiating the immune response (8). In the muscle (**Figure S1A**), the model captures the abstracted hypothesis that MPL and QS-21 activate a ‘muscle-resident immune cell,’ which through chemokine secretion, recruits CCR2⁺ LY6C^{hi} Monocytes from the blood, into the injection site. These monocytes are capable of activation induced by MPL (48), antigen capture and migration into the dLN. This mechanism has been observed using lymphatic cannulation in an ovine (sheep) vaccination model with an AS01-adjuvanted vaccine (49). The potential for Monocytes to infiltrate via HEVs is not captured. Muscle-resident DCs (**Figure S1A**) are also capable of activation by MPL, capturing antigen, and undergoing a maturation process. To capture an abstraction of DC maturation, the key stages are distinguished in the model by an “immature DC,” a “maturing DC,” and a “mature DC.” Immature DCs in the model express TLR4 and IFN γ receptor. Maturing DCs also express CCR7 (50) which mediates migration from tissue to the LN paracortex (51), co-stimulatory molecules (CD80/86), IL-12 receptor and vaccine peptide antigen-MHCII (pMHC) complexes. In addition to the migration of DCs and monocytes into the LN, adjuvant and antigen free-flow from the injection site through the afferent lymphatics (AL) (**Figure S1B**) into the dLN (8). After arrival in the dLN, QS-21 co-localizes with CD169⁺ SCS-M, (**Figure S1D**) and can induce IL-18 secretion from these cells (18). In the paracortex (**Figure S1C**), secreted IL-18 delivers an activation signal to “innate IFN γ secretor cells” (ISC) (10, 25). This cell type is an abstraction to promote

simplicity, encompassing Natural Killer cells, Natural Killer T-cells, innate-like CD8⁺ T-cells, ILC1s, and gamma-delta T-cells, which have all been shown to contribute to the early, innate production of IFN γ after AS01 stimulation (10). IL-18 stimulation of ISCs is capable of promoting the production of IFN γ , augmented by IL-12 (10). During a secondary response, IFN γ levels are further augmented by IL-2 derived from antigen-specific CD4⁺ T cells, further promoting synergistic production of IFN γ . Furthermore, SCS-M (along with follicular dendritic cells) can capture free-flowing antigen, and transfer it to B-cells in the follicle (52), contributing to their priming. The capacity of activated monocytes to differentiate into DCs (8). IL-12 in the model is hypothesized to be secreted by a pool of DCs, [including monocyte-derived-dendritic-cells (MoDCs)] and activated monocytes. At early time points, consistent with literature, IFN γ and IL-12 production is thought to promote differentiation of naïve cognate CD4⁺ T-cells toward a T Helper 1 (Th1) polarization (53). T helper cells are captured in the model, although there is an abstraction at this level for diagrammatic simplicity—the cell is a single entity, where no distinction is made between Th subsets. In the lower-level models, the appropriate distinctions between phenotype of these cells are captured. Here, the Th cell provides expansion, immunoglobulin switching and survival signal to B-cells, or secretes Th1-associated cytokines (TNF- α , IL-2, IFN γ). The model also captures the key stages of antigen-specific B cell priming, activation, and differentiation into “antibody-secreting-cells” or memory B cells, and the formation of germinal centers (**Figure S1E**). The blood compartment of the model (**Figure S1F**) captures the circulation of Th subsets after lymph node egress, and antibody circulation. The remit of the modeling exercise only requires the capture of the generation of T cell and antibody responses, and not the quality or functionality of the antibody response, nor the characteristics of the T-cell response in peripheral tissues, so these are not modeled (for inclusion and exclusion criteria, see **Datasheet 2**). Other inflammatory cytokines such as IL6, IL-1 β etc. are produced during a response to AS01 however it is unclear how they contribute to early immune response to AS01.

AN EXEMPLAR CAPTURING AND DESCRIBING KEY COMPONENTS OF AS01 MOA

To construct a simulation, lower-level behaviors, function, and interactions of components (cells and cytokines) must also be captured, thus substantiating the model. With respect to the research questions, we aim to capture an appropriately detailed description of the biology, building on, and informed by the scope of the CellDesigner model. An adapted version of the unified modeling language (UML) was employed to develop “state-machine” and “activity” diagrams (**Figures 2A,B**) (34, 35). Thus, the Domain Model comprises the CellDesigner visualization and state and activity diagrams. State-machine diagrams describe the different states a component (entity) can exist in, and requirements that govern transitions between

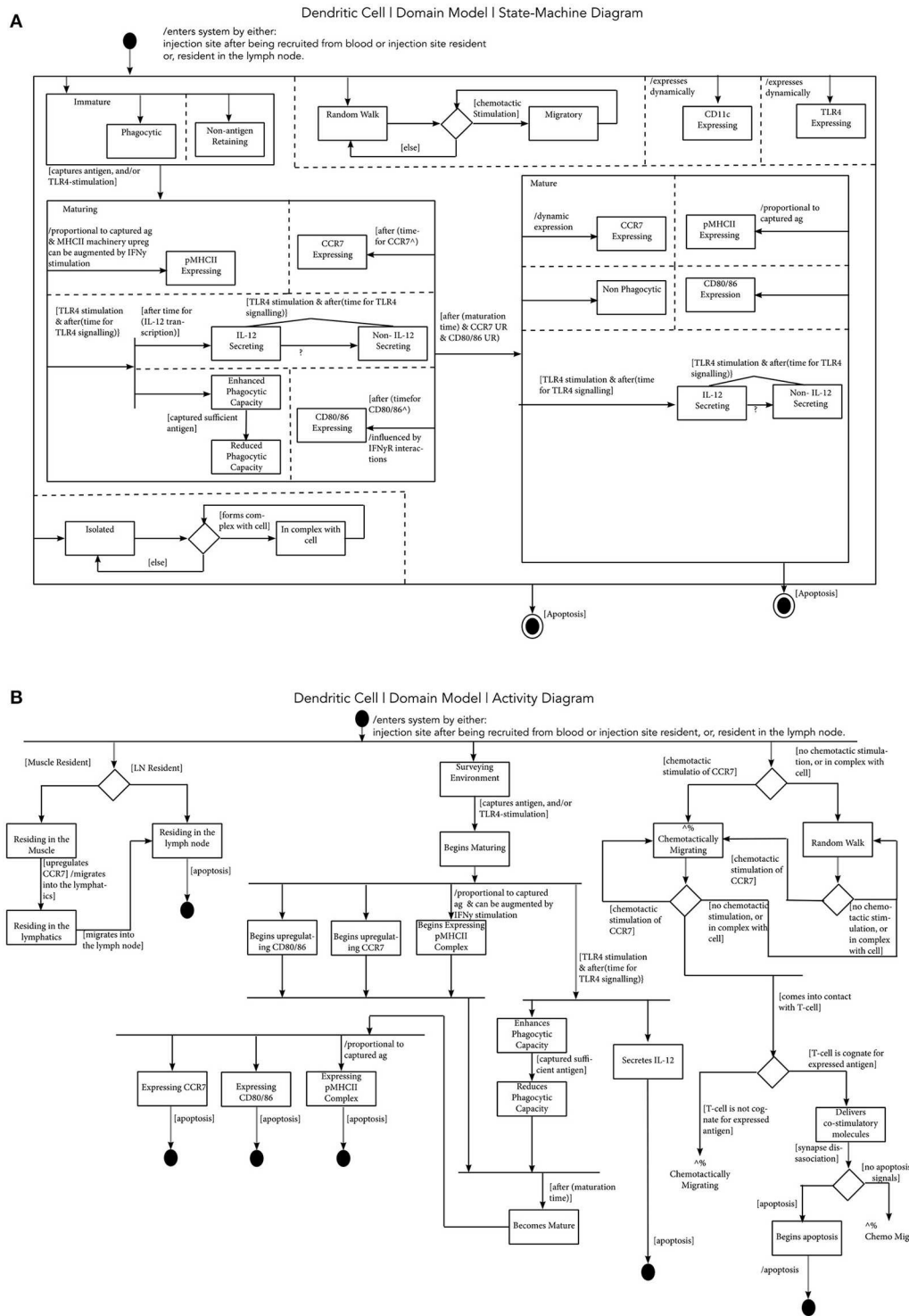


FIGURE 2 | Modeling Dendritic Cell Function. **(A)** A Dendritic Cell State-Machine Diagram: states are represented by rectangular boxes, which may encompass smaller rectangular boxes, indicating parent and sub-states, respectively. A black circle indicates the initial state of the object, and the “double” circle indicates the final state. Arrows indicate transitions between the states, and information surrounded by square brackets describe a condition that must be satisfied for the transition to proceed. An expression preceded by a forward slash, indicates further information with respect to a state transition or a state. Finally, dashed lines represent concurrent states and a diamond indicates a decision is to be made, determined by a condition being satisfied. **(B)** A Dendritic Cell Activity Diagram: in this diagram, a black circle indicates the start of processes, and a double circle indicates the end of activities. The rectangles indicate activities and diamonds represent decisions. Horizontal black lines, are forks and joins, which indicate, respectively, the start and end of activities that occur concurrently.

these states. Activity diagrams in this context describe activities and events in the system that emerge from interactions between cellular components. Care must be taken to capture the relevant biology, and where abstractions are made, these should be appropriate, and decisions behind inclusion and exclusion of functionality and rationale for abstractions should be clearly documented to maintain a trail of the considerations involved, providing transparency of the decisions behind model development in order to permit scrutiny. Our laboratories have developed a tool for this purpose and the reader is directed to Alden et al. (42) for a detailed explanation.

Figure 2A is a state-machine diagram for a murine DC. This captures an abstraction of a cell type, encompassing the different DC subsets in the model (including MoDCs). This diagram captures the key functionality, at the required level of complexity of the DC for our research context. This begins with the cell entering the system, resident in either the injection site or in the dLN. Initially, the cell is immature, and whilst immature, it is phagocytic, and doesn't retain antigen (54). If it captures antigen and/or is stimulated by a TLR4 agonist, it enters the "maturing" state, where the DC has varying phagocytic capacity dependent on levels of antigen it has acquired, increases expression of pMHCII complexes, CCR7, CD80/86, and secretes IL-12. The increased expression of CCR7 allows the cell to chemotactically migrate toward CCL19/21 gradients in both draining lymphatics and in the dLN paracortex. After the time required to undergo full maturation, and if there is sufficient upregulation of CCR7 and CD80/86, the cell can become fully mature. When a DC is fully mature, it is not phagocytic, expresses high levels of pMHCII complexes, CCR7, CD80/86, and can secrete IL-12. During all life stages, this cell dynamically expresses CD11c, and TLR4. Mature DCs can also present antigen to T cells, as the diagram also captures the ability for a DC to be isolated or in a complex with another cell (i.e., when presenting antigen), and to be undergoing random or chemotactic migration. The cell exits the system if it undergoes apoptosis.

Activity diagrams are developed to describe actions and interactions of cellular components. In **Figure 2B** an activity diagram is shown for a murine DC. Initially, the cell is either muscle or dLN resident, surveying the environment, and either undergoing random or chemotactic migration. Following exposure to antigen, and/or stimulation by a TLR-4 agonist, the DC can begin maturing, as described in the state-diagrams, and can eventually become fully mature. If a cell is resident in the muscle site, due to the upregulation of CCR7, it can begin migration into the lymphatics, toward the dLN. If the cell is resident in the dLN, the upregulation of CCR7 would direct it toward the CCL19/21 gradients (which functionally, would direct it to the T cell area) (50). If the DC comes into contact with a T cell, and it undergoes a cognate interaction, the cell can deliver co-stimulatory molecules promoting activation, and either undergo apoptosis or return to migration. Not shown in this perspective, this same approach has been applied to all cell types and processes captured in the AS01 MoA CellDesigner model (see **Datasheet 1** for a list of diagrams).

MODELING AND SIMULATION AS A BASIS FOR EXPLORING ADJUVANT MOA

Following the process outlined in this perspective, simulation can be utilized to integrate knowledge and explore hypotheses underpinning biological systems. The development of Domain Model diagrams can undergo iterative refinement driven by specific scientific questions, resulting in a domain model appropriate to address a specific question. Following this, a Platform Model is developed, and subsequently, a simulation can be constructed written in computer code (44). Simulations are then calibrated to real-world data, and usually undergo a process of validation. After construction, simulations can be inspected by a variety of analysis techniques, such as sensitivity and robustness analysis, permitting an exploration of the effects of stipulated immunological behaviors on the system (41, 45). This can elucidate important MoAs, which can be explored and validated *in vivo* (29, 35), thus guiding experimental design. Furthermore, systems-based techniques that explore optimization could be used to elucidate more efficient dosing schedules. For detailed reading on the entire modeling process described here, the reader is directed to Andrews et al. (44).

CONCLUSION

We have presented a framework to capture and collate MoA knowledge and applied it to integrating and exploring AS01 MoA. This framework has informed the development of a Domain Model, capturing high-level AS01 MoA using CellDesigner, which was further substantiated through UML diagrams describing lower-level functionality. The CellDesigner visualization is an explicit description of the key, higher-level biology, resulting in a visualization with which researchers can illustrate and share their ideas and communicate knowledge and knowledge gaps. Building upon this, the UML-like diagrams captures detailed knowledge and hypotheses underpinning the system, bringing together AS01 understanding, immunological literature, and rational assumption to describe lower-level component behaviors. The resulting Domain Model not only brings together understanding about the biological system, but after appropriate refinement driven by a specific scientific question, can serve as a biological basis to construct simulations, permitting exploration of key research questions. We believe that these concepts will complement work on AS01 MoA and envision that these 3Rs-based approaches (<https://www.nc3rs.org.uk/the-3rs>), through viewing data holistically and complementing *in vivo* experimentation, can be applied more generally to improve the understanding of other adjuvant MoA, thus enhancing vaccine efficacy.

AUTHOR CONTRIBUTIONS

PB, JT, KA, CC, AC, CA, MCC, AD, RM, and MC were involved in the conception and design of the study.

CC, AC, CA, and MC acquired the data. CC, AC, CA, MCC, PB, JT, and KA analyzed the data. PB, AD, RM, ST, MC, JT, KA, CA, and MCC analyzed and interpreted the results. All authors were involved in drafting the manuscript or critically revising it for important intellectual content. All authors had full access to the data and approved the manuscript before it was submitted by the corresponding authors.

FUNDING

Research was funded through a PHD studentship award from BBSRC.

ACKNOWLEDGMENTS

The authors thank Isabelle Carletti for support regarding the data and analysis, Walther Dewé, Tej Patel, Nabila Amanchar, Caroline Hervé (GSK), and Jason Cosgrove (Curie) for fruitful

discussions. This work was partially funded by GlaxoSmithKline Biologicals SA. Shingrix is a trade marks of the GSK group of companies.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.02150/full#supplementary-material>

Figure S1 | The CellDesigner Model captures the following key compartments: the muscle injection site (**A**), the lymphatic vessels (**B**), the draining lymph node (dLN) [including paracortex (**C**), B cell follicles (**D**), germinal center (**E**)], the blood (**F**), and the bone marrow (**G**). This model captures high-level mechanisms that are hypothesized to give rise to the phenomena observed by experimentation in mice. This begins with the intramuscular injection (34) of the vaccine at time zero (**A**) and captures the events leading to the secretion and circulation of antibodies (**E,F**), and lymph node egression of effector memory T cells (**G**).

Datasheet 1 | Full list of Domain Model Diagrams: the full list of diagrams that were created for the AS01 MoA Domain Model.

Datasheet 2 | A list of the entire Core AS01 MoA Domain Model Diagrams.

REFERENCES

- Awate S, Babiuk LA, Mutwiri G. Mechanisms of action of adjuvants. *Front Immun.* (2013) 4:114. doi: 10.3389/fimmu.2013.00114
- Riley EM, Stewart VA. Immune mechanisms in malaria: new insights in vaccine development. *Nat Med.* (2013) 19:168–78. doi: 10.1038/nm.3083
- Morel S, Didierlaurent A, Bourguignon P, Delhay S, Baras B, Jacob V, et al. Adjuvant system AS03 containing α -tocopherol modulates innate immune response and leads to improved adaptive immunity. *Vaccine.* (2011) 29:2461–73. doi: 10.1016/j.vaccine.2011.01.011
- Ciabattini A, Pettini E, Fiorino F, Pastore G, Andersen P, Pozzi G, et al. Modulation of primary immune response by different vaccine adjuvants. *Front Immun.* (2016) 7:427. doi: 10.3389/fimmu.2016.00427
- Reed SG, Orr MT, Fox CB. Key roles of adjuvants in modern vaccines. *Nat Med.* (2013) 19:1597–608. doi: 10.1038/nm.3409
- Alving CR, Beck Z, Matyas GR, Rao M. Liposomal adjuvants for human vaccines. *Expert Opin Drug Deliv.* (2016) 13:5247–1–10. doi: 10.1517/17425247.2016.1151871
- Kool M, Fierens K, Lambrecht BN. Alum adjuvant: some of the tricks of the oldest adjuvant. *J Med Microbiol.* (2012) 61:927–34. doi: 10.1099/jmm.0.038943-0
- Didierlaurent AM, Collignon C, Bourguignon P, Wouters S, Fierens K, Fochesato M, et al. Enhancement of adaptive immunity by the human vaccine adjuvant AS01 depends on activated dendritic cells. *J Immunol.* (2014) 193:1920–30. doi: 10.4049/jimmunol.1400948
- Didierlaurent AM, Laupèze B, Di Pasquale A, Hergli N, Collignon C, Garçon N. Adjuvant system AS01: helping to overcome the challenges of modern vaccines. *Expert Rev Vaccines.* (2016) 16:55–63. doi: 10.1080/14760584.2016.1213632
- Coccia M, Collignon C, Hervé C, Chalon A, Welsby I, Detienne S, et al. Cellular and molecular synergy in AS01-adjuvanted vaccines results in an early IFN γ response promoting vaccine immunogenicity. *NPJ Vaccines.* (2017) 2:25. doi: 10.1038/s41541-017-0027-3
- Cunningham AL, Lal H, Kovac M, Chlibek R, Hwang S-J, Díez-Domingo J, et al. Efficacy of the herpes zoster subunit vaccine in adults 70 years of age or older. *N Engl J Med.* (2016) 375:1019–32. doi: 10.1056/NEJMoa1603800
- James SF, Chahine EB, Sucher AJ, Hanna C. Shingrix: the new adjuvanted recombinant herpes zoster vaccine. *Ann Pharmacother.* (2018) 57:673–80. doi: 10.1177/1060028018758431
- Van Der Meeren O, Hatherill M, Nduba V, Wilkinson RJ, Muyoyeta M, Van Brakel E, et al. Phase 2b controlled trial of M72/AS01 EVaccine to prevent tuberculosis. *N Engl J Med.* (2018) 379:1621–34. doi: 10.1056/NEJMoa1803484
- Alderson MR, McGowan P, Baldrige JR, Probst P. TLR4 agonists as immunomodulatory agents. *J Endotoxin Res.* (2006) 12:313–9. doi: 10.1177/09680519060120050701
- Reed SG, Hsu FC, Carter D, Orr MT. The science of vaccine adjuvants: advances in TLR4 ligand adjuvants. *Curr Opin Immunol.* (2016) 41:85–90. doi: 10.1016/j.coi.2016.06.007
- Didierlaurent AM, Morel S, Lockman L, Giannini SL, Bisteau M, Carlsen H, et al. AS04, an aluminum salt- and TLR4 agonist-based adjuvant system, induces a transient localized innate immune response leading to enhanced adaptive immunity. *J Immunol.* (2009) 183:6186–97. doi: 10.4049/jimmunol.0901474
- Baldwin SL, Roefen W, Singh SK, Tiendrebeogo RW, Christiansen M, Beebe E, et al. Synthetic TLR4 agonists enhance functional antibodies and CD4+ T-cell responses against the Plasmodium falciparum GM22.6C multi-stage vaccine antigen. *Vaccine.* (2016) 34:2207–15. doi: 10.1016/j.vaccine.2016.03.016
- Detienne S, Welsby I, Collignon C, Wouters S, Coccia M, Delhay S, et al. Central role of CD169+ lymph node resident macrophages in the adjuvant activity of the QS-21 component of AS01. *Sci Rep.* (2016) 6:1597. doi: 10.1038/srep39475
- Marty-Roix R, Vladimir GI, Pouliot K, Weng D, Buglione-Corbett R, West K, et al. Identification of QS-21 as an inflammasome-activating molecular component of saponin adjuvants. *J Biol Chem.* (2016) 291:1123–36. doi: 10.1074/jbc.M115.683011
- Okemoto K, Kawasaki K, Hanada K, Miura M, Nishijima M. A potent adjuvant monophosphoryl lipid A triggers various immune responses, but not secretion of IL-1 β or activation of caspase-1. *J Immunol.* (2006) 176:1203–8. doi: 10.4049/jimmunol.176.2.1203
- Leroux-Roels G, Van Belle P, Vandepapeliere P, Horsmans Y, Janssens M, Carletti I, et al. Vaccine adjuvant systems containing monophosphoryl lipid A and QS-21 induce strong humoral and cellular immune responses against hepatitis B surface antigen which persist for at least 4 years after vaccination. *Vaccine.* (2015) 33:1084–91. doi: 10.1016/j.vaccine.2014.10.078
- van den Berg RA, Coccia M, Ballou WR, Kester KE, Ockenhouse CF, Vekemans J, et al. Predicting RTS,S vaccine-mediated protection from transcriptomes in a malaria-challenge clinical trial. *Front Immun.* (2017) 8:557. doi: 10.3389/fimmu.2017.00557

23. Rathinam VA, Fitzgerald KA. Inflammasome complexes: emerging mechanisms and effector functions. *Cell*. (2016) 165:792–800. doi: 10.1016/j.cell.2016.03.046
24. Guo H, Callaway JB, Ting JP. Inflammasomes: mechanism of action, role in disease, and therapeutics. *Nat Med*. (2015) 21:677–87. doi: 10.1038/nm.3893
25. Kastenmüller W, Torabi-Parizi P, Subramanian N, Lämmermann T, Germain RN. A spatially-organized multicellular innate immune response in lymph nodes limits systemic pathogen spread. *Cell*. (2012) 150:1235–48. doi: 10.1016/j.cell.2012.07.021
26. Trewavas A. A brief history of systems biology. “Every object that biology studies is a system of systems.” Francois Jacob. (1974). *Plant Cell*. (2006) 18:2420–30. doi: 10.1105/tpc.106.042267
27. Pappalardo F, Flower D, Russo G, Pennisi M, Motta S. Computational modelling approaches to vaccinology. *Pharmacol Res*. (2015) 92:40–5. doi: 10.1016/j.phrs.2014.08.006
28. Mirsky HP, Miller MJ, Linderman JJ, Kirschner DE. Systems biology approaches for understanding cellular mechanisms of immunity in lymph nodes during infection. *J Theor Biol*. (2011) 287:160–70. doi: 10.1016/j.jtbi.2011.06.037
29. Kirschner D, Pienaar E, Marino S, Linderman JJ. A review of computational and mathematical modeling contributions to our understanding of *Mycobacterium tuberculosis* within-host infection and treatment. *Curr Opin Syst Biol*. (2017) 3:170–85. doi: 10.1016/j.coisb.2017.05.014
30. Alden K, Timmis J, Andrews PS, Veiga-Fernandes H, Coles MC. Pairing experimentation and computational modeling to understand the role of tissue inducer cells in the development of lymphoid organs. *Front Immun*. (2012) 3:172. doi: 10.3389/fimmu.2012.00172
31. Regules JA, Cicatelli SB, Bennett JW, Paolino KM, Twomey PS, Moon JE, et al. Fractional third and fourth dose of RTS,S/AS01 malaria candidate vaccine: a phase 2a controlled human malaria parasite infection and immunogenicity study. *J Infect Dis*. (2016) 214:762–71. doi: 10.1093/infdis/jiw237
32. Wigginton JE, Kirschner DE. A model to predict cell-mediated immune regulatory mechanisms during human infection with *Mycobacterium tuberculosis*. *J Immunol*. (2001) 166:1951–67. doi: 10.4049/jimmunol.166.3.1951
33. Marino S, Myers A, Flynn JL, Kirschner DE. TNF and IL-10 are major factors in modulation of the phagocytic cell environment in lung and lymph node in tuberculosis: A next-generation two-compartmental model. *J Theor Biol*. (2010) 265:586–98. doi: 10.1016/j.jtbi.2010.05.012
34. Cosgrove J, Butler J, Alden K, Read M, Kumar V, Cucurull-Sanchez L, et al. Agent-based modeling in systems pharmacology. *CPT Pharmacometrics Syst Pharmacol*. (2015) 4:615–29. doi: 10.1002/psp4.12018
35. Butler JA, Cosgrove J, Alden K, Timmis J, Coles MC. Model-driven experimentation: a new approach to understand mechanisms of tertiary lymphoid tissue formation, function, and therapeutic resolution. *Front Immun*. (2017) 7:658. doi: 10.3389/fimmu.2016.00658
36. Kisilitsyn A, Savinkov R, Novkovic M, Onder L, Bocharov G. Computational approach to 3D modeling of the lymph node geometry. *Computation*. (2015) 3:222–34. doi: 10.3390/computation3020222
37. Winslow RL, Trayanova N, Geman D, Miller MI. Computational medicine: translating models to clinical care. *Science Transl Med*. (2012) 4:158rv11. doi: 10.1126/scitranslmed.3003528
38. Dada JO, Mendes P. Multi-scale modelling and simulation in systems biology. *Integr Biol (Camb)*. (2011) 3:86–96. doi: 10.1039/c0ib00075b
39. Albergante L, Timmis J, Beattie L, Kaye PM. A petri net model of granulomatous inflammation: implications for IL-10 mediated control of *Leishmania donovani* Infection. *PLoS Comput Biol*. (2013) 9:e1003334. doi: 10.1371/journal.pcbi.1003334
40. Read M, Andrews PS, Timmis J, Williams RA, Greaves RB, Sheng H, et al. Determining disease intervention strategies using spatially resolved simulations. *PLoS ONE*. (2013) 8:e80506. doi: 10.1371/journal.pone.0080506
41. Alden K, Read M, Timmis J, Andrews PS, Veiga-Fernandes H, Coles M. Spartan: a comprehensive tool for understanding uncertainty in simulations of biological systems. *PLoS Comput Biol*. (2013) 9:e1002916. doi: 10.1371/journal.pcbi.1002916
42. Alden K, Andrews PS, Polack FA, Veiga-Fernandes H, Coles MC, Timmis J. Using argument notation to engineer biological simulations with increased confidence. *J R Soc Interface*. (2015) 12:20141059. doi: 10.1098/rsif.2014.1059
43. Read M, Andrews PS, Timmis J, Kumar V. Modelling biological behaviours with the unified modelling language: an immunological case study and critique. *J R Soc Interface*. (2014) 11:20140704. doi: 10.1098/rsif.2014.0704
44. Andrews PS, Polack FAC, Sampson AT, Stepney S, Timmis J. *The CoSMoS Process, Version 0.1: A Process for the Modelling and Simulation of Complex Systems*. University of York Technical Report, YCS-2010-453, Department of Computer Science, University of York (2010) 1–40. Available online at: <http://www-users.cs.york.ac.uk/susan/bib/ss/nonstd/tr453.pdf>
45. Alden K, Andrews PS, Veiga-Fernandes H, Timmis J, Coles M. Utilising a simulation platform to understand the effect of domain model assumptions. *Nat Comput*. (2014) 14:99–107. doi: 10.1007/s11047-014-9428-7
46. Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, et al. The systems biology graphical notation. *Nat Biotechnol*. (2009) 27:735–41. doi: 10.1038/nbt1558
47. Funahashi A, Morohashi M, Kitano H, Tanimura, N. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO*. (2003) 1:159–62. doi: 10.1016/S1478-5382(03)02370-9
48. Michaud JP, Hallé M, Lampron A, Thériault P, Préfontaine P, Filali M, et al. Toll-like receptor 4 stimulation with the detoxified ligand monophosphoryl lipid A improves Alzheimer's disease-related pathology. *Proc Natl Acad Sci USA*. (2013) 110:1941–6. doi: 10.1073/pnas.1215165110
49. Neeland MR, Shi W, Collignon C, Taubenheim N, Meeusen EN, Didierlaurent AM, et al. The lymphatic immune response induced by the adjuvant AS01: a comparison of intramuscular and subcutaneous immunization routes. *J Immunol*. (2016) 197:2704–14. doi: 10.4049/jimmunol.1600817
50. Förster R, Schubel A, Breitfeld D, Kremmer E, Renner-Müller I, Wolf E, et al. CCR7 coordinates the primary immune response by establishing functional microenvironments in secondary lymphoid organs. *Cell*. (1999) 99:23–33. doi: 10.1016/S0092-8674(00)80059-8
51. Martín-Fontecha A, Lanzavecchia A, Sallusto F. Dendritic cell migration to peripheral lymph nodes. *Handb Exp Pharmacol*. (2009) 188:31–49. doi: 10.1007/978-3-540-71029-5_2
52. Junt T, Moseman EA, Iannaccone M, Massberg S, Lang PA, Boes M, et al. Subcapsular sinus macrophages in lymph nodes clear lymph-borne viruses and present them to antiviral B cells. *Nature*. (2007) 450:110–4. doi: 10.1038/nature06287
53. Luckheerem RV, Zhou R, Verma AD, Xia B. CD4 +T cells: differentiation and functions. *Clin Dev Immunol*. (2012) 2012:1–12. doi: 10.1155/2012/925135
54. Lipscomb MF, Masten BJ. Dendritic cells: immune regulators in health and disease. *Physiol Rev*. (2002) 82:97–130. doi: 10.1152/physrev.00023.2001

Conflict of Interest Statement: CC, AC, CA, RM, AD, ST, and MC are employees of the GSK group of companies. CA, RM, AD, MC, and ST report ownership of GSK shares and/or restricted GSK shares. MC was supported by a Marie Skłodowska Curie Intra-European Fellowship (ref. “ADJSYN”). PB was holding a Ph.D. studentship and collaborated with GSK at the time of the study as part of his Ph.D. training. PB's work was partially funded by GlaxoSmithKline Biologicals SA through a post-graduate studentship.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Buckley, Alden, Coccia, Chalon, Collignon, Temmerman, Didierlaurent, van der Most, Timmis, Andersen and Coles. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership