Immunology and Cell Biology (2011) 89, 81–89

www.nature.com/icb

REVIEW

Innovative bioinformatic approaches for developing peptide-based vaccines against hypervariable viruses

Danylo Sirskyj[1], Francisco Diaz-Mitoma[1], Ashkan Golshani[2], Ashok Kumar[3,4] and Ali Azizi[1,3,4]

The application of the fields of pharmacogenomics and pharmacogenetics to vaccine design has been recently labeled 'vaccinomics'. This newly named area of vaccine research, heavily intertwined with bioinformatics, seems to be leading the charge in developing novel vaccines for currently unmet medical needs against hypervariable viruses such as human immunodeficiency virus (HIV), hepatitis C and emerging avian and swine influenza. Some of the more recent bioinformatic approaches in the area of vaccine research include the use of epitope determination and prediction algorithms for exploring the use of peptide epitopes as vaccine immunogens. This paper briefly discusses and explores some current uses of bioinformatics in vaccine design toward the pursuit of peptide vaccines for hypervariable viruses. The various informatics and vaccine design strategies attempted by other groups toward hypervariable viruses will also be briefly examined, along with the strategy used by our group in the design and synthesis of peptide immunogens for candidate HIV and influenza vaccines.
Immunology and Cell Biology (2011) **89,** 81–89; doi:10.1038/icb.2010.65; published online 11 May 2010

Keywords: vaccinomics; peptide vaccines; epitope mapping; HIV; hypervariable viruses

The application of the fields of pharmacogenomics and pharmacogenetics to vaccine design, heavily integrated with bioinformatics, has been recently termed 'vaccinomics'.[1] The integration of bioinformatics with vaccinology has likely arisen for several reasons. The laborious and imperfect nature of conventional experimentally based approaches can be easily seen as a bottleneck toward developing new vaccines. Conventional methods rely on antigens expressed in sufficient amounts from *in vitro* culture models. With this approach exists the possibility that transiently expressed proteins may be missed, leading to the possibility of interesting candidate antigens being overlooked. In addition, abundantly produced antigens may not necessarily make strong vaccine candidates.[2] The rapid emergence of new viral pathogens, such as the recent avian and swine influenza strains, underscores the need for improved and expedited processes for developing and producing vaccines in response to such outbreaks. Timely responses to such outbreaks simply cannot be enacted by conventional approaches, and thus current methods may benefit from a more rapid, *in silico* informatics-based approach. One of the key drivers of such an informatics-based approach has been the recent sequencing of many pathogen genomes, as well as the increase in nucleotide and protein sequence databases.[2] To complement this surge in data, new bioinformatics programs, such as T- and B-cell epitope mapping programs (reviewed in De Groot and Berzofsky[3] and De Groot[4]), have been developed to properly use the abundance of data being generated. Advances in bioinformatics applications have also resulted in the appearance of programs such as motif-based systems,[5] support vector machines,[6–8] hidden Markov models,[9] neural networks,[10] quantitative structure–activity relationship analyses[11] and structure-based approaches[12,13] geared toward predicting major histocompatibility complex (MHC)-peptide binding. It should be emphasized that, ultimately, such predictive *in silico* work needs to be confirmed through *in vivo* experiments to fully validate such approaches.

VIRAL PEPTIDE IMMUNOGENS: EPITOPE DETERMINATION, DESIGN AND ALGORITHMS

Thus far, most commercialized antiviral vaccines have been derived from inactivated or live attenuated viruses or recombinant proteins. Although they have been successful so far against a range of viral pathogens, the effectiveness of current vaccines against hypervariable viruses is limited. Heterogeneity among hypervariable viruses, such as human immunodeficiency virus (HIV) and hepatitis C virus (HCV), poses a significant challenge to vaccine development.[14–17] For instance, HIV-1 has eight subtypes, with a high degree of diversity even within each subtype.[18] Such hypervariable RNA viruses evade the immune mechanisms of host species because of the high mutation rate of their genomes, which comes as a result of transcriptional errors.[19] The immune system cannot catch up with the rate of change and the viruses escape recognition and control, resulting in disease or death.

Thus, the urgent need for improved vaccines capable of addressing such impasses is obvious. Indeed, various approaches are being undertaken in the development of such improved vaccines. Among the various strategies employed are the use of reverse vaccinology

[1]Variation Biotechnologies Inc., Ottawa, Ontario, Canada; [2]Department of Biology, Carleton University, Ottawa, Ontario, Canada; [3]Infectious Disease and Vaccine Research Center, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada and [4]Department of Pathology and Laboratory Medicine, University of Ottawa, Ottawa, Ontario, Canada
Correspondence: Dr A Azizi, Department of Pathology and Laboratory Medicine, University of Ottawa, 451 Smyth, Ottawa, Ontario, Canada K1H 8M5.
E-mail: aazizi@uottawa.ca
Received 14 December 2009; revised 22 March 2010; accepted 11 April 2010; published online 11 May 2010

approaches and the rational attenuation of live viruses. Reverse vaccinology involves the use of computational methods to identify all potential candidate immunogens from within the genomic sequence of a pathogen.[2,20,21] Once appropriate vaccine candidates have been identified from the pathogen genome, genes of interest are then cloned and used to produce their corresponding protein. *In vivo* and *in vitro* testing and characterization is subsequently performed on the cloned proteins of interest to further validate their potential use as vaccine immunogens. Indeed, such an approach is significantly quicker than conventional culture-based approaches, and has had success in elucidating vaccine candidates for various bacterial pathogens.[2,20] Attenuated live viral vaccines have been shown to induce more robust responses over alternative vaccine types.[2,22] The process of attenuating live viruses for use in vaccines has previously involved laboriously passaging the virus at progressively lower temperatures.[23] A recent approach by Lim *et al*.[24] describes a computational approach to engineering attenuated vesicular stomatitis viruses. Their model was able to successfully predict experimentally observed effects of gene re-arrangements on the growth characteristics of vesicular stomatitis virus mutants.

An alternative approach involves the use of synthetic polyvalent peptide vaccines representing various T- and B-cell epitopes. The use of polyvalent epitopes, representing the multiple possible variants of a virus corresponding to diverse human leukocyte antigen (HLA) types in each population, may help induce protection against multiple strains of a particular virus.[25–29] This strategy requires an understanding of T- and B-cell epitopes in the microorganism's proteins and their interaction with MHC or HLA complexes. A successful peptide-based vaccine must also include immunodominant T- and B-cell epitopes;[30,31] that is, epitopes that are readily accessed and recognized by the immune system and that have influence on the specificity of the induced antibody. The idea of using such peptide epitopes was conceived from the scrutiny of hundreds of overlapping synthetic peptides. This analysis revealed that only a small number of regions in a protein are immunogenic and capable of provoking humoral and cellular immune responses. For example, B cells recognize epitopes exposed on the surface of antigens, whereas T cells distinguish specific amino acid sequences that are first recognized by MHC class I and II molecules on the surface of antigen-presenting cells.[32–34] Compared with other vaccine types, synthetic peptide vaccines offer many advantages as well as some disadvantages. Advantages include an absence of any potentially infectious material, such as that present in live attenuated vaccines; the ability to include multiple epitopes, accounting for multiple variants of a pathogen; economical production on a large scale; and the ability to manufacture vaccines for pathogens difficult or impossible to culture in the laboratory. The main benefit of vaccination with peptides would be the ability to minimize the amount and complexity of a well-defined antigen. An appropriate peptide-based vaccine would also decrease the chance of stimulating a response against self-antigens, thus providing a safer vaccine by preventing the induction of autoimmunity. Some pitfalls in using peptide vaccines include the poor immunogenicity of short peptides, requiring the use of potent adjuvants, and the need to tailor administered epitopes with regard to the polymorphic structure of HLA.[29] Peptide dimerization and aggregation or insolubility at physiologic pH ranges are other challenges that must be addressed in the development of peptide-based vaccines.[26] Recent developments in the field of peptide therapeutics have seen the creation of small molecules capable of accelerating the rate of peptide binding and display by MHC molecules on the surface of dendritic cells. The functionality of these compounds was shown both *in vitro* and *in vivo*,[35] and may allow for the vastly improved delivery of peptide therapeutics and vaccines.

The main driving force behind synthetic peptide vaccines has been the development of novel bioinformatics applications, namely T-cell and B-cell epitope determination programs, as well as HLA-binding prediction algorithms, transporter of antigen processing (TAP) affinity prediction algorithms and proteasomal cleavage prediction algorithms. Examples of such programs are shown in Table 1. Combining these programs in one sequential analysis allows for the scanning of pathogenic genomes for open reading frames and the further identification of potentially immunogenic epitopes and their propensity for TAP and HLA binding. Such epitopes may then be synthesized and screened *in vitro* by humoral and cellular assays such as enzyme-linked immunosorbent assay, enzyme-linked immunosorbent spot or intracellular cytokine staining to evaluate the immunogenic potential of the selected peptides. Effective epitopes, or their parental protein, may then be considered as potential vaccine candidates.[3] Preclinical studies of peptide-based vaccines have indicated that the routes of administration, dose and type of adjuvant all have a crucial role in determining the efficacy of such vaccines.

T-CELL EPITOPE MAPPING AND PREDICTION

As described above, a successful peptide-based vaccine must include immunodominant epitopes.[30,31] One of the problems facing traditional vaccines is the lack of a broad cell-mediated immune response against variable pathogens.[36–39] Humoral immunity may help prevent infection, but to date only a limited number of antibodies with neutralizing capability have been identified for viruses such as HIV and HCV,[40–43] and predicaments in the induction of neutralizing antibodies against such diverse viral variants have encouraged scientists to focus on cell-mediated immune responses.[44,45] The induction of cell-mediated immune responses with a large repertoire of immune specificities has emerged as an essential characteristic for the clearance or control of hypervariable viral infections such as HCV and HIV.[46–49] In an impressive study, Jones *et al*.[50] showed the presence of a high frequency of CD8+ T cells in HIV-resistant sex workers in Nairobi.

The question arises as to whether a peptide-based vaccine is able to provoke broad cellular immunity against such hypervariable viruses. To prepare more effective immunogens, some studies have shown that peptide epitope-based vaccines should contain epitopes capable of inducing T-helper responses along with cytotoxic T-cell responses or antibody responses.[28] Thus, continued efforts toward the elucidation of immunodominant T-cell epitopes for various viral pathogens remains critical.

Given the importance of T-cell responses in controlling viral infections, the larger number of T-cell epitope mapping and prediction algorithms available today comes as no surprise.[3,4] One of the more comprehensive programs seems to be EpiMatrix from EpiVax Inc. (http://www.epivax.com/). The EpiMatrix tool set is able to predict epitopes against over 100 different MHC class I and class II alleles. In a typical EpiMatrix analysis the target protein sequence is broken down into overlapping 9-mer frames in which each frame overlaps the last by eight amino acids. Each of the derived 9-mer frames is then screened for predicted affinity against a panel of MHC class I and/or class II alleles. The resulting scores fall on a common scale that can be directly compared across HLA alleles. The ability to rate putative epitopes on a common scale is described as an exclusive feature of the EpiMatrix system. The EpiMatrix platform is also closely tied with additional computational tools such as ClustiMer (scans EpiMatrix results for T-cell epitope 'clusters'), BlastiMer (automated BLAST search tool), OptiMatrix (involved in deimmunizing

Table 1 Algorithms and databases for bioinformatics-driven vaccine research

Tools and Databases	*Description*
PAP http://bio.dfci.harvard.edu/Tools/antigenic	This program predicts segments from within a protein sequence that are likely to be antigenic by eliciting an antibody response
EPIMHC http://bio.dfci.harvard.edu/epimhc/	A curated database of MHC ligands
PEPVAC http://immunax.dfci.harvard.edu/PEPVAC/	This program fully covers multi-epitope vaccines based on genomewide predictions of MHC class I epitopes
BCIPEP http://www.imtech.res.in/raghava/bcipep/index.html	This database has a collection of the peptides having a role in humoral immunity
IEDB http://www.immuneepitope.org/	This program contains data related to antibody and T-cell epitopes
RANKPEP http://bio.dfci.harvard.edu/Tools/rankpep.html	This program predicts peptide binders to MHC-I and MHC-II molecules
SYFPEITHI http://www.syfpeithi.de/	A database containing thousands of peptide sequences known to bind class I and II MHC molecules
NetMHC 3.0 http://www.cbs.dtu.dk/services/NetMHC/	This program predicts binding of peptides to a number of different HLA alleles
DiscoTope 1.2 Server http://www.cbs.dtu.dk/services/DiscoTope/	This program predicts discontinuous B-cell epitopes from three-dimensional protein structures
NetCTL 1.2 Server http://www.cbs.dtu.dk/services/NetCTL/	It predicts CTL epitopes in protein sequences
CTL PRED http://www.imtech.res.in/raghava/ctlpred/	It predicts CTL epitopes that are crucial in vaccine design
HIV Molecular Immunology Database http://www.hiv.lanl.gov/content/immunology/index.html	This database searches collection of HIV-1 cytotoxic and helper T-cell epitopes and antibody-binding sites
Influenza Research Database http://www.fludb.org/brc/home.do?Decorator=influenza	This database provides resources to facilitate the development of vaccines, diagnostics and therapeutics for these pathogens
HCV sequence database http://hcv.lanl.gov/content/sequence/HCV/ToolsOutline.html	This database searches collection of HCV cytotoxic and helper T-cell epitopes and antibody-binding sites
MHCBN http://www.imtech.res.in/raghava/mhcbn/	A database of MHC binding and non-binding peptides
PREDEPP http://margalit.huji.ac.il/	MHC class-I epitope prediction software
EpiMatrix http://www.epivax.com/	EpiVax's commercial epitope prediction platform
NetChop 3.1 server http://www.cbs.dtu.dk/services/NetChop/	Neural network prediction for cleavage sites of the human proteasome
PAProC http://www.paproc.de/	Prediction tool for cleavage by human and yeast 20S proteasomes based on experimental cleavage data
Epibase http://www.algonomics.com/immunogenicity/epibase_overview.php	Algonomics commercial T-cell epitope screening tool. Predicts epitopes of proteins with known structures

Abbreviations: CTL, cytotoxic T lymphocyte; HCV, hepatitis C virus; HIV, human immunodeficiency virus; HLA, human leukocyte antigen; MHC, major histocompatibility complex.

sequences), Conservatrix (involved in finding conserved epitopes) and Vaccine CAD (an *in silico* vaccine design algorithm).[51]

Algonomics (Gent, Belgium) is another player in the T-cell epitope screening realm. Algonomics' Epibase platform predicts and screens for T-cell epitopes, and using a structure-based approach, it is described as being capable of predicting peptide-MHC II interactions from various HLA allotypes from Caucasian, Oriental, Hispanic and Afro-American population groups. Epibase may be used for both epitope determination and de-immunization applications of protein therapeutics.[52]

Despite the continual advances in T-cell epitope prediction algorithms, such approaches have yet to be perfected. In a study examining naturally processed and HLA class I-presented peptides from vaccinia virus infection, Johnson *et al.*[53] observed that computer algorithms failed to predict 20% of the peptides determined using a mass spectrometry-based approach. This was thought to be because of limitations in the length of the output sequence produced by the predictive algorithms. Such limitations of computer-based predictive algorithms may result in potentially important epitope regions being missed.

B-CELL EPITOPE MAPPING AND PREDICTION

Although there exist multiple T-cell epitope mapping algorithms, B-cell epitope mapping algorithms are far fewer, owing to the difficulty in being able to determine B-cell epitopes in humans. B-cell epitopes are defined by a specific surface region of an antigenic protein, and may be divided into two different types of epitopes: continuous (linear) epitopes and discontinuous (conformational) epitopes.[54] Thus, conformational context is much more critical in B-cell epitope prediction than it is for T-cell epitope prediction, and consequently, knowledge of conformational epitopes effective at eliciting neutralizing antibodies would be greatly beneficial in vaccine design. Examples of B-cell epitope-mapping algorithms include 3DEX (3D-Epitope-Explorer),[55] CEP (conformational epitope predictor)[56] and DiscoTope.[57] 3DEX software is designed to allow the localization of linear peptide sequences within the three-dimensional structure of a protein. The program achieves this by using an algorithm that takes into account the physicochemical properties of Cα- or Cβ- atoms of the individual amino acids within the sequence. The software also integrates a 'joker' function, which allows the allocation of peptide sequences in the absence of complete homology between the primary protein sequence and the tertiary protein structure. A self-described caveat of this software is that it lacks a final analysis step involving energy minimizations of the mapped sequence, in comparison to other methods;[58] however, critical factors such as surface exposure and spacial relationships between neighboring amino acids are taken into consideration. CEP predicts epitopes of proteins with known structures using accessibility of residues and spatial distance cutoffs to predict antigenic determinants, conformational epitopes and sequential epitopes.[56,59,60] Another B-cell epitope predictor, DiscoTope, was designed specifically for the prediction of discontinuous B-cell epitopes.[57] This predictor takes into account amino acid statistics, spatial information and surface accessibility for the prediction of B-cell epitopes. In addition, this prediction method was trained using a data set of discontinuous epitopes determined from X-ray crystallographic structures of antibody-antigen protein complexes.[57] In a test of its

abilities, DiscoTope was shown to successfully confirm discontinuous epitope residues that were mapped by alternative methods such as phage-display and point mutation. The inclusion of such structural data within this method is thought to be responsible for the improved performance of this method, and indeed these findings were in agreement with those observed elsewhere.[61]

Despite the availability of such prediction algorithms, such tools seem to be only experimental tools at best. Previous work by Blythe and Flower[62] has shown amino acid propensity scales to be ineffective in predicting linear B-cell epitopes, and subsequent evaluation of antibody and protein binding site prediction web servers by Ponomarenko and Bourne,[63] when compared with X-ray crystallographic methods, indicated overall poor performance of all tools evaluated. In both instances, evaluation by the researchers suggested that the majority of the methods examined performed close to random in their ability to predict epitopes. In 2007, NIAID (National Institute of Allergy and Infectious Diseases) investigators attending a workshop to evaluate the state of B-cell epitope prediction programs found that the current state of B-cell epitope prediction tools was far from ideal, and proposed several recommendations for improvements in this area, such as implementing high-quality data sets for training and testing novel prediction algorithms, adopting the use of a common metric to evaluate predictor tool performance and adopting standardization of data formats as well as for program inputs and outputs.[64] In 2008, Sweredoski and Baldi[65] published the release of PEPITO, a discontinuous B-cell epitope predictor, using a combination of amino acid propensity scores and side chain orientation and solvent accessibility information using half sphere exposure values, incorporating some of the recommendations of the 2007 NIAID workshop. A recent publication by Dominguez *et al.*,[67] using a B-cell epitope prediction algorithm devised by Kolaskar and Tongaonkar,[66] showed some success in predicting conserved neutralization-relevant B-cell epitopes on the surface of the merozoite parasite *Babesia bovis*, although only 2/4 epitopes showed effective neutralization of the parasite *in vitro*.[67] Such results encouragingly suggest that B-cell epitope prediction methods may also be effectively used in the development of epitope or subunit-based veterinary vaccines, though further work is required to improve epitope prediction.

Characterization of strongly neutralizing epitopes of hypervariable viruses remains elusive. Although many mouse influenza B-cell epitopes are known, only one known human epitope is known for influenza A.[68] For instance, only a few neutralizing human monoclonal antibodies such as 2F5 and 4E10 are found with HIV-1 core-binding specificities. In the approach used by our group, hypervariable regions of viral surface proteins were located and analyzed through proprietary bioinformatic software. The amino acid sequences of viral proteins from hundreds of infected individuals, of various HLA types, were aligned and the location of these variable sequences was transposed onto the protein's three-dimensional structure by molecular modeling, paying particular attention to hypervariable epitopes exposed on the surface of the virus, easily accessed by B-cell or T-cell receptors. An identical approach was also used by our group in developing an epitope-based candidate influenza vaccine. Predicted hypervariable B-cell epitopes were then synthesized, and the potency of these B-cell epitopes for inducing humoral responses was evaluated using immunoglobulin G (IgG) enzyme-linked immunosorbent assay and a competitive neutralization assay measuring antibody binding to the synthesized peptides using sera from infected, convalescent or immunized individuals. Peptides that were able to bind antibodies from patient sera were then further evaluated for their immunogenic potential *in vivo* in mice, in combination with appropriate adjuvants. The *in vivo* potency of these predicted B-cell epitopes was subsequently further measured using IgG and IgA enzyme-linked immunosorbent assay, microneutralization assay,[69] intracellular cytokine staining and B-cell enzyme-linked immunosorbent spot assay. Similar to what has been observed in the literature regarding the effectiveness of B-cell prediction algorithms, not all of our synthesized epitopes were shown to be immunogenic. The approach regarding our candidate HIV vaccine is discussed in more detail later in this paper.

Overall, it would seem that B-cell epitope prediction algorithms, although improving, clearly require further work, and indeed the NIAID workshop conducted in 2007 provides some guidance in this area. Although no epitope-based vaccines are currently licensed for use in humans, it remains to be observed which B-cell epitope prediction approach shall be first to help reach this milestone.

BIOINFORMATICS APPROACHES TO PREDICT ANTIGEN–ANTIBODY INTERACTIONS

In recent years, significant advances have been made in the computational prediction of protein–protein interactions. Given that experimental techniques to design and detect interactions between proteins remain resource and time demanding, along with the technical limitations associated with them, improved computational tools represent important progress in the field. Computational tools can be used to predict novel interacting partners, to assess and validate the presumed interactions and to help design novel vaccines.

Bioinformatics tools that are used to predict protein–protein interactions can be classified into five general categories: methods based on genomic context, evolutionary relationships, three-dimensional structure of proteins, presence of specific protein domains and the primary structure of proteins. These methods use the information from experimentally determined interactions to predict novel ones. Genomic-based approaches take advantage of the availability of genome sequence information and the information about what genes are present and in what physical order. Conservation of gene orders across different species is used as a basis for a functional correlation and hence a physical interaction.[70] Tools that are based on evolutionary relationships take advantage of the fact that protein pairs with similar phylogenetic profiles (co-conservation) in different genomes are thought to interact.[71] Three-dimensional protein structures can also be used to predict docking domains of two potentially interacting proteins. They may be used to predict such interactions as between peptide-loaded MHC molecules with T-cell receptors. The compatibility of the interacting regions of proteins are used to predict the likelihood of such an interaction.[72] The presence of interacting domains can also be used to identify interacting partners.[73] In this case, the query proteins are searched against the database of interacting domains for the presence of regions with similarity to domains that are known to mediate physical interactions.[74] Lastly, methods have also been developed to recognize sequence patterns from primary protein structures to predict interactions.[75]

One of the more recently developed bioinformatics tools uses the co-occurrence of short polypeptide sequences ($\leqslant$25 amino acids) to predict a physical interaction between proteins from their primary structures.[76] This approach, which is called protein interaction prediction engine, or PIPE, is based on the assumption that a finite number of short polypeptide codes mediate the interactions between some proteins. These codes are assumed to be significantly shorter than the classical domains and are used repeatedly in different proteins and in different contexts within the cell. The probability of an

interaction is estimated by the number of co-occurrences for the corresponding codes in the data set of experimentally determined interacting proteins. Recently, the specificity of this approach was increased to >99.95%.[77] In this way, a number of short polypeptide codes were identified, which in theory could mediate an interaction between two otherwise non-interacting proteins. The identification of novel short polypeptide codes that can mediate protein–protein interactions may have a significantly impact on the design of novel vaccines. In agreement with this, it was recently shown that addition of a short polypeptide code to a non-interacting protein mediated its novel interaction with a target protein.[78] With respect to the field of vaccines, in particular mucosally administered vaccines, such protein–protein interactions may have a role in improving the delivery of mucosally or orally administered vaccines. Intestinal microfold (M) cells are specialized epithelial cells that predominantly reside in the follicle-associated epithelium overlying Peyer's patches, and uptake antigens or microorganisms from the intestinal lumen by phago-, endo- or pinocytosis and transcytosis, and deliver them to the underlying immune system of the mucosae.[79–81] Mechanisms of pathogen entry into M cells may be exploited in developing new mucosal vaccines. To date, only limited numbers of M-cell receptors and their ligands have been identified, and most of these receptors are not only expressed in M cells but also in neighboring enterocytes. Several studies have previously shown that *Ulex europaeus* agglutinin-1 (UEA-1), a specific lectin for α-l-fucose residues, selectively binds to mouse Peyer's patches and targets mouse M cells. In a study by Manocha *et al.*,[82] the UEA-1 coated on the surface of microparticles encoding HIV genes had the capability to bind to the apical surface of M cells. However, same as for PRPs, UEA-1 might bind not only to M cells but also to goblet cells and mucous layers.[83] Microarray and three-dimensional imaging of specific molecules associated with M cells has revealed that a surface marker called glycoprotein 2 is expressed in both human and mouse M cells.[84,85] It seems that glycoprotein 2 has an important role in molecular mechanisms responsible for antigen uptake by M cells. It serves as a transcytotic apical receptor on the surface of M cells that specifically bind to type I pili on bacterial outer membranes (FimH).[84] Elimination of glycoprotein 2 reduced the entry and uptake of bacteria into Peyer's patches and decreased T-cell proliferative and antibody responses. Thus, targeting such specific receptors on the apical surface of M cells, perhaps through the inclusion of short polypeptide codes, may have the ability to specifically increase the uptake and presentation of antigens, consequently initiating the immune response and inducing protection against infectious challenge.

APPROACHES TOWARD EPITOPE-BASED VACCINE DESIGN

Because of the existence of multiple epitope mapping and binding prediction algorithms, many groups have applied their bioinformatic approaches to the design of candidate vaccines. We have chosen to examine the efforts of some of the more comprehensively described approaches found in the literature. The efforts of several groups, including our own, are described below.

Previously, the approach used by Epimmune Inc., originally based in San Diego, California, involved the bioinformatic sequence analysis of proteins, with a focus on identifying the presence of specific anchor residue motifs involved in binding peptides to most major HLA class I and II molecules.[86] Affinities between prospective peptides and HLA molecules were determined by use of a matrix that estimated the effects of all amino acids along all possible positions in the peptide chain. Such an approach has been previously described by this group.[87] Together with affinity prediction, an emphasis was placed on conserved regions of pathogens, in an attempt to target viral regions not readily amenable to mutation. Quantitative *in vitro* binding assays were then used to screen for peptides that were sufficiently immunogenic. To address the issue of the highly polymorpic locus of human HLA, Epimmune adopted the identification of peptides capable of binding multiple HLA alleles. When applied to HIV-1 sequences from the Los Alamos database, in the case of cytotoxic T lymphocyte epitopes, overall population covered was estimated at 98% using 17 epitopes. Immunogenicity of their chosen epitopes was further confirmed through *in vivo* studies using HLA transgenic mice and sera samples from pathogen-exposed individuals. In the case of their efforts on a candidate HIV, HBV, HCV and malaria vaccines, 85–100% of their supertype epitopes tested were found to elicit responses from immunized or exposed patients. Additional work from Epimmune Inc. has been published;[88–90] however, it has since been acquired by VaxOnco, a South Korean company focused on the development and commercialization of novel cancer therapies and immunotherapies against tumors and infectious diseases.

EpiVax Inc. has previously applied their software platform toward the design of cross-subtype HIV immunogens.,[91] Their informatics-based approach consisted of identifying conserved and promiscuous class I and class II HLA-restricted T-cell epitopes from a database of 18 313 HIV-1 env sequences to synthesize a pool of epitopes that were both immunogenic and highly conserved across the greatest number of HIV-1 strains possible. Such a strategy, which was focused on stimulating the cellular arm of the immune system, was used in light of the data supporting the importance of the cellular immune response in controlling early HIV infection,[46–49] as has been observed with patients described as long-term nonprogressors.[92–94]

A representative study conducted by De Groot *et al.*[91] began with 18 313 GenBank HIV env sequences available from the Los Alamos National Laboratory HIV database during the year 2000. Excluded were sequences deposited before January 1998 and those shorter than 60–80% of the length of the env component being studied. Next, they used an in-house algorithm termed Conservatrix, which was used to identify protein or DNA segments highly conserved between the input sequences. The output of these conserved regions was limited to 1000 entries, or all peptides conserved in at least 5% of the input sequences, whichever was the least. The conserved peptide regions were then scanned using EpiMatrix, an algorithm used to model interactions between peptides (typically 9-mers) and MHC molecules, and subsequently scored according to several weighted factors, among them an EpiMatrix Z-score with respect to the HLA allele DRB1*0101 (the emphasis for this allele was because of the availability of a transgenic mouse strain subsequently used for *in vivo* experiments, containing the human HLA DRB1*0101). From this list of ranked peptides, immunogenic consensus sequence epitopes were created using EpiVax's EpiAssembler tool. These peptides consisted of a 9-mer core peptide, with seven residues flanking the core region on both sides, for a final 23-residue peptide. These extended peptides were then ranked again, and evaluated for the presence of published ligands from an internal database. From this list, candidate peptides were selected for further *in vivo* testing. The top candidate peptides from above were then BLAST searched against the protein database in GenBank, using the EpiVax BlastiMer algorithm to search for peptides showing significant sequence homology (⩾75%) to human or common viral and bacterial genomes, and any such peptides were removed. Of these final peptides, nine were then synthesized for subsequent *in vivo* testing. Of these final nine peptides, eight were shown to be immunogenic *in vitro*, as evaluated by enzyme-linked immunosorbent spot assays using peripheral blood mononuclear cells from HIV-infected

patients, whereas three contained sequences having been previously identified and published in the literature as being immunogenic.

Another example of a recent bioinformatics-led approach toward vaccine design undertaken by Vider-Shalit *et al.*[95] pertains to the epitope-based design of a vaccine against hepatitis C, with a focus on MHC-I CD8 epitopes. In this example, the genomes of four different subtypes of hepatitis C (1a, 1b, 2a and 3a) were scanned, with a focus on 10 specific genes (core, envelope 1 and 2 (E1 and E2), and non-structural proteins p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B). Each genome sequence was then divided into all possible nonamers (11 754 total). Each of these nonamers was then scored for its ability to undergo (1) proteosomal cleavage, (2) binding to TAP machinery and (3) MHC-I binding (for 29 different class I HLA molecules, limited to the HLA family B*2703). The resulting sequences were next evaluated for the presence of conserved regions. In this example, peptides $<90\%$ conserved were removed from analysis, as the researchers were only interested in conserved regions. Following this, the peptide sequences were evaluated for their homology to self-peptides against 33 869 human protein sequences available in the Ensembl database. In this situation, none of the epitopes showed any similarity to self-epitopes. Lastly, the epitopes that passed the screen against self were subjected to a genetic algorithm to determine a final optimized group of sequences. Subsequent computational processing led to the determination of a final HCV vaccine sequence length of 25 amino acids, representative of the conserved regions of the four subtypes of HCV examined, for the ten genes in question.

Peters *et al.*[96] recently examined mutations in HIV-*gag* sequences from HIV-positive Kenyan women using various bioinformatic methods. The purpose of their informatics searches was to identify positively selected amino acids involved in the mutational escape of HIV-1 from controlling cytotoxic T lymphocyte responses, and to evaluate the potential effects of such mutations on various stages of antigen processing, to infer how the virus will mutate and identify any vulnerabilities this may expose. In brief, their approach involved examining proviral *gag* sequences and phylogenetically classifying the viral subtypes using MEGA (Molecular Evolutionary Genetic Analysis) software. A selection-mapping algorithm (QUASI) was used to identify positively selected amino acids (mutations that were overabundant compared with silent mutations at each codon), and NetChop C-term 3.0 was used to determine the cleavage likelihood of *gag* residues. This was followed by a TAP affinity prediction step to evaluate the propensity of the mutated *gag* sequences for TAP transport efficiency. Ultimately, integrated predictions of C-terminal proteasomal cleavage, TAP affinity and HLA class I associations were correlated with $CD4^+$ counts obtained from the study patients. Although this bioinformatics approach may have yielded designs for potentially effective immunogens, no *in vivo* study was conducted at the time to confirm their observations.

As has been previously described,[97] using TAP prediction algorithms as a filter before predicting peptide–MHC-I binding affinities may improve the quality of the output sequences. In the above examples by Vider-Shalit *et al.*[95] and Peters *et al.*,[96] including TAP prediction algorithms may serve to improve the final output; however, we believe that such predictive methods ultimately require more comprehensive *in vivo* follow-up. Although undoubtedly useful, it remains to be observed whether such TAP prediction methods ultimately find their way into a method that is able to predict epitopes capable of inducing functional and protective immune responses.

An immense deal of interest has been focused on the development of candidate vaccines targeting conserved regions of HIV-1; however, none of these approaches were able to induce broadly crossreactive immunity *in vivo*.[98–100] Through a bioinformatic approach, our group devised a new technology called Variosite technology. Using this approach, we have developed two highly effective, synthetic peptide-based vaccines. In brief, using proprietary software, hundreds of HIV-1 and influenza genome sequences were analyzed to design peptide immunogens that represented the antigenic diversity in variable epitopes. Once the extent of antigenic variation within a chosen epitope has been determined, Variosite technology uses algorithms to determine at which positions within the epitope variability should be represented in the immunogen design. Although some amino acid positions within an epitope are quite variable, limits exist in their variation. Our peptide immunogens are considerably longer than those used in the past (>20 amino acids in length, rather than 9–10 amino acids), which promotes generation of a beneficial T-helper cell response, and second, our peptides target highly variable immunogenic regions of the viral proteins rather than conserved regions, which may have greater immunological value in the form of a polyvalent vaccine. Once designed, many antigenic variants of a given epitope are synthesized simultaneously using a method that is robust and reproducible. A similar such approach taking into account HIV sequence diversity has also been attempted by Frahm *et al.*[101]

Below is a description of the synthesis of a cocktail of peptide variants, collectively termed a 'Variosite' immunogen or Variosite preparation. The Variosite-based technology is an improvement of a precursor technology called hypervariable epitope constructs,[102] which allows us to represent the extensive antigenic diversity in these regions.

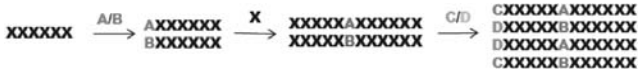
The sequence of the variable epitope of interest =**XXXXXVXXXXX** (where **X** represents a well-conserved amino acid in the sequence and **V** represents a position with variable numbers of amino acids;

If at position **V** a mixture of desired amino acids (say **A and B**) is used instead of a single amino acid, the following mixture of peptides will be simultaneously synthesized:

XXXXXAXXXXX

XXXXXBXXXXX

By extension, having two variable positions along the peptide chain and placing a mixture of two amino acids at those positions, we get four different peptide analogs during a single synthesis run:



To advance our proof of concept, we have recently published details and shown the immunogenicity in non-human primates of our prototype, Variosite-based HIV-1 vaccine.[103] This multivalent HIV-1 candidate vaccine comprised a pool of 176 peptides representing variable regions of HIV-1; 5 hypervariable epitopes in the gp120 envelope, and 2 variable epitopes in the gag protein. The potency and breadth of this polyvalent vaccine approach was tested against a panel of heterologous HIV-1 subtypes in a non-human primate model. Specific $CD8^+$ T-cell immune responses against HIV-1 subtypes A–F were shown. These results were noteworthy as HIV-1 sequences within a subtype differ by up to 20%, and between subtypes by up to 35%. In addition, binding antibody titer and neutralizing activity was characterized in immunized animals. A substantial level of IgG antibody titer to variant gp120 proteins was observed in all animals. Neutralizing antibody was also measured against T-cell line-adapted and primary isolates of HIV-1 subtypes. Interestingly, three out of six immunized macaques developed neutralization activity against two primary HIV-1

isolates from subtype D. To our knowledge, this is the first time that neutralizing antibodies were shown against HIV-1 isolates using linear hypervariable epitopes as immunogens. Our unpublished data on influenza-based Variosite formulations have also shown the ability of our vaccine designs to protect against drifted strains of influenza virus, emphasizing that the approach can represent past, present and anticipated antigenic variability in the epitopes targeted by our vaccine. In addition, our efforts with external vendors have shown the feasibility of gram-scale production of good manufacturing practice (GMP)-quality synthetic peptide immunogens, with the possibility of scaling up to the kg scale. This shows that such an approach can be readily applied to large-scale, GMP-quality production.

CLINICAL TRIALS

Despite the occurrence of many peptide-based vaccine studies in animals, within the realm of infectious diseases, only a limited number of peptide vaccines have been assessed in human clinical trials. In a phase 1 placebo-controlled trial, the efficacy of a multiepitope HIV-1 peptide vaccine along with granulocyte macrophage colony-stimulating factor as an adjuvant was evaluated. Although the vaccine did not elicit any adverse effects, only 6 of 80 volunteers showed moderate HIV-specific immune responses.[104] In another phase I study, 33 HIV-seronegative volunteers were primed orally three times with a V3 peptide derived from the HIV-1 isolate MN, followed by a systemic boosting. However, none of the individuals showed broad humoral or cellular immune responses in mucosal sites.[105] In a recent human clinical trial, an HCV core protein (C35-44, YLLPRRGPRL) was injected into infected individuals of various HLA types. After six vaccine injections, specific cytotoxic T lymphocyte activity was shown to be augmented in 15 of 25 patients against the injected peptides. In addition, specific IgG production was augmented in 15 of 22 patients after 12 vaccine administrations. Furthermore, two patients showed a 1 log viral load reduction after treatment.[106] Another recent peptide vaccine trial examined the use of the M2e epitope, conserved in both human and avian influenza A viruses. The peptide alone showed low immunogenicity; however, fusing M2e to hepatitis B virus-derived virus-like particles as a carrier showed promising results. In animal models, this chimeric vaccine administered parenterally or intranasally showed protection against various influenza A strains. In human trials, this vaccine was shown to be both safe and immunogenic against influenza.[107]

Overall, these clinical trials have shown that this approach is safe, cost effective and relatively effective in eliciting peptide-specific immune responses. However, to validate the protective efficacy induced by such vaccines, further human clinical trials are required.

CONCLUDING REMARKS

Given the pervasiveness of viruses such as HIV and HCV, and often sudden appearance of infectious diseases such as SARS (severe acute respiratory syndrome) and avian and swine influenza, it is easily observed that improved and more rapidly dispensible vaccines are urgently needed. The problems and limitations of conventional influenza vaccine manufacturing processes have been recently underscored in 2009 during the H1N1 influenza pandemic. Similarly for HIV and HCV, an effective prophylactic vaccine remains elusive despite extensive efforts put forward over the past 20 years. The emergence of these infectious diseases has made it very clear that conventional vaccine approaches are incapable of addressing the immunologic issues created by such hypervariable viruses, and new approaches and technologies are urgently needed. The close integration of the fields of vaccinomics with bioinformatics has initiated a new era of *in silico*-driven peptide epitope vaccine design. Advances in this emerging field may lead to significantly more effective vaccines against hypervariable viruses. To continue progressing forward, existing informatics databases should be continually improved and integrated with existing and newly developed analytical tools, enabling researchers to analyze greater numbers of parameters in vaccine development. Ultimately, such computational efforts need to be validated *in vivo* to evaluate the effectiveness of the immunogens predicted *in silico*.

ACKNOWLEDGEMENTS

AA thanks the Ontario HIV Treatment Network (Grant ROGB G116) for support.

1 Poland GA, Ovsyannikova IG, Jacobson RM. Application of pharmacogenomics to vaccines. *Pharmacogenomics* 2009; **10**: 837–852.
2 Flower DR. *Bioinformatics for Vaccinology*. John Wiley and Sons, West Sussex, UK, 2008.
3 De Groot AS, Berzofsky JA. From genome to vaccine—new immunoinformatics tools for vaccine design. *Methods* 2004; **34**: 425–428.
4 De Groot AS. Immunomics: discovering new targets for vaccines and therapeutics. *Drug Discov Today* 2006; **11**: 203–209.
5 Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 1999; **50**: 213–219.
6 Donnes P, Elofsson A. Prediction of MHC class I binding peptides, using SVMHC. *BMC Bioinformatics* 2002; **3**: 25.
7 Liu W, Meng X, Xu Q, Flower DR, Li T. Quantitative prediction of mouse class I MHC peptide binding affinity using support vector machine regression (SVR) models. *BMC Bioinformatics* 2006; **7**: 182.
8 Wan J, Liu W, Xu Q, Ren Y, Flower DR, Li T. SVRMHC prediction server for MHC-binding peptides. *BMC Bioinformatics* 2006; **7**: 463.
9 Noguchi H, Kato R, Hanai T, Matsubara Y, Honda H, Brusic V *et al*. Hidden Markov model-based prediction of antigenic peptides that interact with MHC class II molecules. *J Biosci Bioeng* 2002; **94**: 264–270.
10 Burden FR, Winkler DA. Predictive Bayesian neural network models of MHC class II peptide binding. *J Mol Graph Model* 2005; **23**: 481–489.
11 Reche PA, Reinherz EL. PEPVAC: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands. *Nucleic Acids Res* 2005; **33**: W138–W142.
12 Davies MN, Hattotuwagama CK, Moss DS, Drew MG, Flower DR. Statistical deconvolution of enthalpic energetic contributions to MHC-peptide binding affinity. *BMC Struct Biol* 2006; **6**: 5.
13 Doytchinova IA, Walshe V, Borrow P, Flower DR. Towards the chemometric dissection of peptide—HLA-A*0201 binding affinity: comparison of local and global QSAR models. *J Comput Aided Mol Des* 2005; **19**: 203–212.
14 Azizi A, Anderson DE, Ghorbani M, Gee K, Diaz-Mitoma F. Immunogenicity of a polyvalent HIV-1 candidate vaccine based on fourteen wild type gp120 proteins in golden hamsters. *BMC Immunol* 2006; **7**: 25.
15 Gaschen B, Taylor J, Yusim K, Foley B, Gao F, Lang D *et al*. Diversity considerations in HIV-1 vaccine selection. *Science* 2002; **296**: 2354–2360.
16 Naas T, Ghorbani M, varez-Maya I, Lapner M, Kothary R, De Repentigny Y *et al*. Characterization of liver histopathology in a transgenic mouse model expressing genotype 1a hepatitis C virus core and envelope proteins 1 and 2. *J Gen Virol* 2005; **86**: 2185–2196.
17 Onafuwa-Nuga A, Telesnitsky A. The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol Mol Biol Rev* 2009; **73**: 451–480, Table.
18 Gao F, Robertson DL, Carruthers CD, Li Y, Bailes E, Kostrikis LG *et al*. An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). *J Virol* 1998; **72**: 10234–10241.
19 Fiers W, De FM, Birkett A, Neirynck S, Min JW. A 'universal' human influenza A vaccine. *Virus Res* 2004; **103**: 173–176.
20 Masignani V, Rappuoli R, Pizza M. Reverse vaccinology: a genome-based approach for vaccine development. *Expert Opin Biol Ther* 2002; **2**: 895–905.
21 Rappuoli R, Covacci A. Reverse vaccinology and genomics. *Science* 2003; **302**: 602.
22 Chen GL, Subbarao K. Live attenuated vaccines for pandemic influenza. *Curr Top Microbiol Immunol* 2009; **333**: 109–132.
23 Murphy BR, Coelingh K. Principles underlying the development and use of live attenuated cold-adapted influenza A and B virus vaccines. *Viral Immunol* 2002; **15**: 295–323.
24 Lim KI, Lang T, Lam V, Yin J. Model-based design of growth-attenuated viruses. *PLoS Comput Biol* 2006; **2**: e116.
25 Arnon R, Ben-Yedidia T. Old and new vaccine approaches. *Int Immunopharmacol* 2003; **3**: 1195–1204.

26 Azizi A, Diaz-Mitoma F. Viral peptide immunogens: current challenges and opportunities. *J Pept Sci* 2007; **13**: 776–786.
27 Kawada M, Igarashi H, Takeda A, Tsukamoto T, Yamamoto H, Dohki S *et al.* Involvement of multiple epitope-specific cytotoxic T-lymphocyte responses in vaccine-based control of simian immunodeficiency virus replication in rhesus macaques. *J Virol* 2006; **80**: 1949–1958.
28 Purcell AW, Zeng W, Mifsud NA, Ely LK, Macdonald WA, Jackson DC. Dissecting the role of peptides in the immune response: theory, practice and the application to vaccine design. *J Pept Sci* 2003; **9**: 255–281.
29 Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov* 2007; **6**: 404–414.
30 Agadjanyan MG, Ghochikyan A, Petrushina I, Vasilevko V, Movsesyan N, Mkrtichyan M *et al.* Prototype Alzheimer's disease vaccine using the immunodominant B cell epitope from beta-amyloid and promiscuous T cell epitope pan HLA DR-binding peptide. *J Immunol* 2005; **174**: 1580–1586.
31 Udhayakumar V, Anyona D, Kariuki S, Shi YP, Bloland PB, Branch OH *et al.* Identification of T and B cell epitopes recognized by humans in the C-terminal 42-kDa domain of the plasmodium falciparum merozoite surface protein (MSP)-1. *J Immunol* 1995; **154**: 6022–6030.
32 Jung MC, Diepolder HM, Pape GR. T cell recognition of hepatitis B and C viral antigens. *Eur J Clin Invest* 1994; **24**: 641–650.
33 Lu NQ. A speculative view of immune recognition. *Immunol Invest* 1994; **23**: 53–71.
34 Takahashi H. Antigen presentation in vaccine development. *Comp Immunol Microbiol Infect Dis* 2003; **26**: 309–328.
35 Call MJ, Xing X, Cuny GD, Seth NP, Altmann DM, Fugger L *et al. In vivo* enhancement of peptide display by MHC class II molecules with small molecule catalysts of peptide exchange. *J Immunol* 2009; **182**: 6342–6352.
36 Carruth LM, Greten TF, Murray CE, Castro MG, Crone SN, Pavlat W *et al.* An algorithm for evaluating human cytotoxic T lymphocyte responses to candidate AIDS vaccines. *AIDS Res Hum Retroviruses* 1999; **15**: 1021–1034.
37 Nabel GJ. HIV vaccine strategies. *Vaccine* 2002; **20**: 1945–1947.
38 Niedrig M, Gregersen JP, Fultz PN, Broker M, Mehdi S, Hilfenhaus J. Immune response of chimpanzees after immunization with the inactivated whole immunodeficiency virus (HIV-1), three different adjuvants and challenge. *Vaccine* 1993; **11**: 67–74.
39 Yasutomi Y, Palker TJ, Gardner MB, Haynes BF, Letvin NL. Synthetic peptide in mineral oil adjuvant elicits simian immunodeficiency virus-specific CD8+ cytotoxic T lymphocytes in rhesus monkeys. *J Immunol* 1993; **151**: 5096–5105.
40 Li M, Gao F, Mascola JR, Stamatatos L, Polonis VR, Koutsoukos M *et al.* Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol* 2005; **79**: 10108–10125.
41 Locher CP, Grant RM, Collisson EA, Reyes-Terán G, Elbeik T, Kahn JO *et al.* Antibody and cellular immune responses in breakthrough infection subjects after HIV type 1 glycoprotein 120 vaccination. *AIDS Res Hum Retroviruses* 1999; **15**: 1685–1689.
42 Sarmati L, d'Ettorre G, Nicastri E, Ercoli L, Uccella I, Massetti P *et al.* Neutralizing antibodies against autologous human immunodeficiency virus type 1 isolates in patients with increasing CD4 cell counts despite incomplete virus suppression during antiretroviral treatment. *Clin Diagn Lab Immunol* 2001; **8**: 822–824.
43 Srivastava IK, Ulmer JB, Barnett SW. Neutralizing antibody responses to HIV: role in protective immunity and challenges for vaccine design. *Expert Rev Vaccines* 2004; **3**: S33–S52.
44 Duerr A, Wasserheit JN, Corey L. HIV vaccines: new frontiers in vaccine development. *Clin Infect Dis* 2006; **43**: 500–511.
45 Luckay A, Sidhu MK, Kjeken R, Megati S, Chong SY, Roopchand V *et al.* Effect of plasmid DNA vaccine design and *in vivo* electroporation on the resulting vaccine-specific immune responses in rhesus macaques. *J Virol* 2007; **81**: 5257–5269.
46 Azizi A, Ghorbani M, Kryworuchko M, Aucoin S, Diaz-Mitoma F. Potency of cell-mediated immune responses to different combined HIV-1 immunogens in a humanized murine model. *Hum Vaccin* 2005; **1**: 170–176.
47 Azizi A, Ghorbani M, Soare C, Mojibian M, Diaz-Mitoma F. Synergistic effect of combined HIV/HCV immunogens: a combined HIV-1/HCV candidate vaccine induces a higher level of CD8+ T cell-immune responses in HLA-A2.1 mice. *Curr HIV Res* 2007; **5**: 211–219.
48 Cristillo AD, Wang S, Caskey MS, Unangst T, Hocker L, He L *et al.* Preclinical evaluation of cellular immune responses elicited by a polyvalent DNA prime/protein boost HIV-1 vaccine. *Virology* 2006; **346**: 151–168.
49 Letvin NL, Mascola JR, Sun Y, Gorgone DA, Buzby AP, Xu L *et al.* Preserved CD4+ central memory T cells and survival in vaccinated SIV-challenged monkeys. *Science* 2006; **312**: 1530–1533.
50 Rowland-Jones SL, Dong T, Fowke KR, Fowke KR, Kimani J, Krausa P *et al.* Cytotoxic T cell responses to multiple conserved HIV epitopes in HIV-resistant prostitutes in Nairobi. *J Clin Invest* 1998; **102**: 1758–1765.
51 EpiVax Inc Technical Platform (http://www epivax.com/platform/). 2009.
52 Algonomics 2009.
53 Johnson KL, Ovsyannikova IG, Mason CJ, Bergen III HR, Poland GA. Discovery of naturally processed and HLA-presented class I peptides from vaccinia virus infection using mass spectrometry for vaccine development. *Vaccine* 2009; **28**: 38–47.
54 Davies MN, Flower DR. Harnessing bioinformatics to discover new vaccines. *Drug Discov Today* 2007; **12**: 389–395.
55 Schreiber A, Humbert M, Benz A, Dietrich U. 3D-Epitope-Explorer (3DEX): localization of conformational epitopes within three-dimensional structures of proteins. *J Comput Chem* 2005; **26**: 879–887.
56 Kulkarni-Kale U, Bhosle S, Kolaskar AS. CEP: a conformational epitope prediction server. *Nucleic Acids Res* 2005; **33**: W168–W171.
57 Haste AP, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 2006; **15**: 2558–2567.
58 Pizzi E, Cortese R, Tramontano A. Mapping epitopes on protein surfaces. *Biopolymers* 1995; **36**: 675–680.
59 Kolaskar AS, Kulkarni-Kale U. Prediction of three-dimensional structure and mapping of conformational epitopes of envelope glycoprotein of Japanese encephalitis virus. *Virology* 1999; **261**: 31–42.
60 Kulkarni-Kale U. *Prediction of 3D Structure and Function of Proteins and Peptides*. University of Pune: India, 2003.
61 Batori V, Friis EP, Nielsen H, Roggen EL. An in silico method using an epitope motif database for predicting the location of antigenic determinants on proteins in a structural context. *J Mol Recognit* 2006; **19**: 21–29.
62 Blythe MJ, Flower DR. Benchmarking B cell epitope prediction: underperformance of existing methods. *Protein Sci* 2005; **14**: 246–248.
63 Ponomarenko JV, Bourne PE. Antibody-protein interactions: benchmark datasets and prediction tools evaluation. *BMC Struct Biol* 2007; **7**: 64.
64 Greenbaum JA, Andersen PH, Blythe M, Bui HH, Cachau RE, Crowe J *et al.* Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools. *J Mol Recognit* 2007; **20**: 75–82.
65 Sweredoski MJ, Baldi P. PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics* 2008; **24**: 1459–1460.
66 Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 1990; **276**: 172–174.
67 Dominguez M, Echaide I, Echaide ST, Mosqueda J, Cetrá B, Suarez CE *et al.* In silico predicted conserved B-cell epitopes in the merozoite surface antigen-2 family of B. bovis are neutralization sensitive. *Vet Parasitol* 2010; **167**: 216–226.
68 Bui HH, Peters B, Assarsson E, Mbawuike I, Sette A. Ab and T cell epitopes of influenza A virus, knowledge and opportunities. *Proc Natl Acad Sci USA* 2007; **104**: 246–251.
69 Sirskyj D, Weltzin R, Golshani A, Anderson D, Bozic J, Diaz-Mitoma F *et al.* Detection of influenza A and B neutralizing antibodies in vaccinated ferrets and macaques using specific biotin-streptavidin conjugated antibodies. *J Virol Methods* 2010; **163**: 459–464.
70 Dandekar T, Snel B, Huynen M, Bork P. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci* 1998; **23**: 324–328.
71 Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci USA* 1999; **96**: 4285–4288.
72 Lu L, Lu H, Skolnick J. MULTIPROSPECTOR: an algorithm for the prediction of protein-protein interactions by multimeric threading. *Proteins* 2002; **49**: 350–364.
73 Han DS, Kim HS, Jang WH, Lee SD, Suh JK. PreSPI: a domain combination based prediction system for protein-protein interaction. *Nucleic Acids Res* 2004; **32**: 6312–6320.
74 Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. The database of interacting proteins: 2004 update. *Nucleic Acids Res* 2004; **32**: D449–D451.
75 Shen J, Zhang J, Luo X, Zhu W, Yu K, Chen K *et al.* Predicting protein-protein interactions based only on sequences information. *Proc Natl Acad Sci USA* 2007; **104**: 4337–4341.
76 Pitre S, Dehne F, Chan A, Cheetham J, Duong A, Emili A *et al.* PIPE: a protein-protein interaction prediction engine based on the re-occurring short polypeptide sequences between known interacting protein pairs. *BMC Bioinformatics* 2006; **7**: 365.
77 Pitre S, North C, Alamgir M, Jessulat M, Chan A, Luo X *et al.* Global investigation of protein-protein interactions in yeast saccharomyces cerevisiae using re-occurring short polypeptide sequences. *Nucleic Acids Res* 2008; **36**: 4286–4294.
78 Neduva V, Linding R, Su-Angrand I, Stark A, de Masi F, Gibson TJ *et al.* Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol* 2005; **3**: e405.
79 Yuki Y, Kiyono H. Mucosal vaccines: novel advances in technology and delivery. *Expert Rev Vaccines* 2009; **8**: 1083–1097.
80 Brayden DJ, Jepson MA, Baird AW. Keynote review: intestinal Peyer's patch M cells and oral vaccine targeting. *Drug Discov Today* 2005; **10**: 1145–1157.
81 Holmgren J, Czerkinsky C. Mucosal immunity and vaccines. *Nat Med* 2005; **11**: S45–S53.
82 Manocha M, Pal PC, Chitralekha KT, Thomas BE, Tripathi V, Gupta SD *et al.* Enhanced mucosal and systemic immune response with intranasal immunization of mice with HIV peptides entrapped in PLG microparticles in combination with Ulex Europaeus-I lectin as M cell target. *Vaccine* 2005; **23**: 5599–5617.
83 Giannasca PJ, Giannasca KT, Leichtner AM, Neutra MR. Human intestinal M cells display the sialyl Lewis A antigen. *Infect Immun* 1999; **67**: 946–953.
84 Terahara K, Yoshida M, Igarashi O, Nochi T, Pontes GS, Hase K *et al.* Comprehensive gene expression profiling of Peyer's patch M cells, villous M-like cells, and intestinal epithelial cells. *J Immunol* 2008; **180**: 7840–7846.
85 Hase K, Kawano K, Nochi T, Pontes GS, Fukuda S, Ebisawa M *et al.* Uptake through glycoprotein 2 of FimH(+) bacteria by M cells initiates mucosal immune response. *Nature* 2009; **462**: 226–230.
86 Sette A, Livingston B, McKinney D, Appella E, Fikes J, Sidney J *et al.* The development of multi-epitope vaccines: epitope identification, vaccine design and clinical evaluation. *Biologicals* 2001; **29**: 271–276.
87 Sette A, Buus S, Appella E, Smith JA, Chesnut R, Miles C *et al.* Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis. *Proc Natl Acad Sci USA* 1989; **86**: 3296–3300.

88 Alexander J, del Guercio MF, Frame B, Maewal A, Sette A, Nahm MH *et al.* Development of experimental carbohydrate-conjugate vaccines composed of Streptococcus pneumoniae capsular polysaccharides and the universal helper T-lymphocyte epitope (PADRE). *Vaccine* 2004; **22**: 2362–2367.
89 Livingston B, Crimi C, Newman M, Higashimoto Y, Appella E, Sidney J *et al.* A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. *J Immunol* 2002; **168**: 5499–5506.
90 McKinney DM, Skvoretz R, Livingston BD, Wilson CC, Anders M, Chesnut RW *et al.* Recognition of variant HIV-1 epitopes from diverse viral subtypes by vaccine-induced CTL. *J Immunol* 2004; **173**: 1941–1950.
91 De Groot AS, Bishop EA, Khan B, Lally M, Marcon L, Franco J *et al.* Engineering immunogenic consensus T helper epitopes for a cross-clade HIV vaccine. *Methods* 2004; **34**: 476–487.
92 Harrer T, Harrer E, Kalams SA, Elbeik T, Staprans SI, Feinberg MB *et al.* Strong cytotoxic T cell and weak neutralizing antibody responses in a subset of persons with stable nonprogressing HIV type 1 infection. *AIDS Res Hum Retroviruses* 1996; **12**: 585–592.
93 Harrer T, Harrer E, Kalams SA, Barbosa P, Trocha A, Johnson RP *et al.* Cytotoxic T lymphocytes in asymptomatic long-term nonprogressing HIV-1 infection. Breadth and specificity of the response and relation to *in vivo* viral quasispecies in a person with prolonged infection and low viral load. *J Immunol* 1996; **156**: 2616–2623.
94 Klein MR, van Baalen CA, Holwerda AM, Kerkhof Garde SR, Bende RJ, Keet IP *et al.* Kinetics of Gag-specific cytotoxic T lymphocyte responses during the clinical course of HIV-1 infection: a longitudinal analysis of rapid progressors and long-term asymptomatics. *J Exp Med* 1995; **181**: 1365–1372.
95 Vider-Shalit T, Raffaeli S, Louzoun Y. Virus-epitope vaccine design: informatic matching the HLA-I polymorphism to the virus genome. *Mol Immunol* 2007; **44**: 1253–1261.
96 Peters HO, Mendoza MG, Capina RE, Luo M, Mao X, Gubbins M *et al.* An integrative bioinformatic approach for studying escape mutations in human immunodeficiency virus type 1 gag in the Pumwani Sex Worker Cohort. *J Virol* 2008; **82**: 1980–1992.
97 Peters B, Bulik S, Tampe R, Van Endert PM, Holzhutter HG. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J Immunol* 2003; **171**: 1741–1749.
98 Braibant M, Brunet S, Costagliola D, Rouzioux C, Agut H, Katinger H *et al.* Antibodies to conserved epitopes of the HIV-1 envelope in sera from long-term non-progressors: prevalence and association with neutralizing activity. *AIDS* 2006; **20**: 1923–1930.
99 Kothe DL, Li Y, Decker JM, Bibollet-Ruche F, Zammit KP, Salazar MG *et al.* Ancestral and consensus envelope immunogens for HIV-1 subtype C. *Virology* 2006; **352**: 438–449.
100 Nyambi PN, Mbah HA, Burda S, Williams C, Gorny MK, Nádas A *et al.* Conserved and exposed epitopes on intact, native, primary human immunodeficiency virus type 1 virions of group M. *J Virol* 2000; **74**: 7096–7107.
101 Frahm N, Kaufmann DE, Yusim K, Muldoon M, Kesmir C, Linde CH *et al.* Increased sequence diversity coverage improves detection of HIV-specific T cell responses. *J Immunol* 2007; **179**: 6638–6650.
102 Meyer D, Anderson DE, Gardner MB, Torres JV. Hypervariable epitope constructs representing variability in envelope glycoprotein of SIV induce a broad humoral immune response in rabbits and rhesus macaques. *AIDS Res Hum Retroviruses* 1998; **14**: 751–760.
103 Azizi A, Anderson DE, Torres JV, Ogrel A, Ghorbani M, Soare C *et al.* Induction of broad cross-subtype-specific HIV-1 immune responses by a novel multivalent HIV-1 peptide vaccine in cynomolgus macaques. *J Immunol* 2008; **180**: 2174–2186.
104 Spearman P, Kalams S, Elizaga M, Metch B, Chiu YL, Allen M *et al.* Safety and immunogenicity of a CTL multiepitope peptide vaccine for HIV with or without GM-CSF in a phase I trial. *Vaccine* 2009; **27**: 243–249.
105 Lambert JS, Keefer M, Mulligan MJ, Schwartz D, Mestecky J, Weinhold K *et al.* A phase I safety and immunogenicity trial of UBI microparticulate monovalent HIV-1 MN oral peptide immunogen with parenteral boost in HIV-1 seronegative human subjects. *Vaccine* 2001; **19**: 3033–3042.
106 Yutani S, Komatsu N, Shichijo S, Yoshida K, Takedatsu H, Itou M *et al.* Phase I clinical study of a peptide vaccination for hepatitis C virus-infected patients with different human leukocyte antigen-class I-A alleles. *Cancer Sci* 2009; **100**: 1935–1942.
107 Fiers W, De FM, El BK, Schepens B, Roose K, Schotsaert M *et al.* M2e-based universal influenza A vaccine. *Vaccine* 2009; **27**: 6280–6283.