Vaccine 33 (2015) 7408–7414

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Current methods of epitope identification for cancer vaccine design

Gregory A. Cherryholmes*, Sasha E. Stanton, Mary L. Disis

Tumor Vaccine Group, Center for Translational Medicine in Women's Health, University of Washington, 850 Republican Street, Seattle, WA 98109, United States

ARTICLE INFO

Article history:
Available online 1 August 2015

Keywords:
Vaccine
Epitope detection
Sequence-based prediction
Structure-based prediction
Self-protein
Peptide
In silico prediction

ABSTRACT

The importance of the immune system in tumor development and progression has been emerging in many cancers. Previous cancer vaccines have not shown long-term clinical benefit possibly because were not designed to avoid eliciting regulatory T-cell responses that inhibit the anti-tumor immune response. This review will examine different methods of identifying epitopes derived from tumor associated antigens suitable for immunization and the steps used to design and validate peptide epitopes to improve efficacy of anti-tumor peptide-based vaccines. Focusing on *in silico* prediction algorithms, we survey the advantages and disadvantages of current cancer vaccine prediction tools.

1. Introduction

The interaction of the immune system in cancer growth and development has long been established; tumors evade detection of the immune system by secreting immunosuppressive cytokines, inhibiting cytotoxic T cell activation by cell–cell signaling and inducing exhaustion, and utilizing angiogenic signaling to build tumor vasculature. However, increased tumor immune infiltrate, particularly Th1 immune infiltrate, predicts improved clinical response in many tumor subtypes [1]. This suggests that identifying ways to modify the immune environment with vaccines may provide significant clinical benefit. Most vaccines currently in clinical trials rely on tumor cell lysate or isolated full length proteins which can be immune suppressive. Using *in silico*-based methods of epitope prediction to produce rationally-designed peptide vaccines could increase efficacy of peptide-based vaccines.

2. Identification of tumor associated antigens for vaccine design

One active area of investigation is identification of the best antigen targets; either tumor associated over-expressed self-proteins or mutated tumor proteins. Many antigen targets were identified based on presence of an immune response (typically auto-antibodies) to self-proteins uniquely over-expressed in tumor-bearing patients. For example, the antigen TERT is overexpressed in almost all cancers, making it a universal tumor antigen (predicted to be associated with over 85% of all cancers) [2]. A phase II trial of administration of a TERT vaccine with gemcitabine and 5 fluorouracil in 73 patients with surgically resected pancreatic cancer showed 62% 1 year disease free survival [3,4] and currently a Phase III trial is ongoing [5]. Unfortunately, TERT is also highly expressed in hematopoietic stem cells leading to concerns over the effect of vaccination with TERT on normal cells [2]. Similarly GP2, a vaccine targeting HER2 overexpressing breast cancer tumors has entered phase III trials for high-risk breast cancer patients after being shown to increase GP2 specific CD8 T cells and epitope spreading in all patients ($P < 0.001$) in a phase I trial of 18 metastatic HER2 overexpressing breast cancer patients [6]. However, HER2 is only expressed by a subset (20–25%) of breast cancers as well as a smaller subset of other cancers including gastric cancer [6–8]. Cancer testis antigens (CTAs) are interesting targets because they are not normally expressed outside the testis and placenta [9]. Currently, four CTAs targeted in late stage Phase II+ clinical trials are NY-ESO-1, MAGE-A3, and LAGE-1. MAGE-A3 is currently being studied for treatment in melanoma, lung, multiple myeloma [10,11]. LAGE-1 and NY-ESO-1 are being used in combination to

Abbreviations: MHC, major histocompatibility complex; pMHC, peptide–MHC complex; CTA, cancer testis antigen; QM, quantitative matrix; DS-QM, docking simulation-quantitative matrix; ANN, artificial neural network; SVM, support vector machine; HMM, hidden markov model; ACS, ant colony strategies; IEDB, internet epitope database; SDR, specific determining residue; QSAR, quantitative structure activity relationship; HPLC, high performance liquid chromatography.

* Corresponding author. Tel.: +1 206 543 8646.
E-mail addresses: gcherryh@medicine.washington.edu (G.A. Cherryholmes), sestant2@medicine.washington.edu (S.E. Stanton), ndisis@medicine.washington.edu (M.L. Disis).

http://dx.doi.org/10.1016/j.vaccine.2015.06.116
0264-410X/

treat multiple myeloma [12]. In the phase II trial of MAGE-A3 in completely resected stage Ib to IIa non-small cell lung cancer, of the 182 patients treated there was no statistically significant difference in disease free (HR 0.76 95% CI 0.48–1.21 $P = 0.248$) or overall survival (HR 0.81 95% CI 0.47–1.40 $P = 0.454$) despite measurable immune response of 119 of 122 patients receiving the MAGE-A3 vaccine as compared to no measurable immune response in the control patients [13]. In another clinical trial, 56 melanoma patients were treated with a MAGE-A3 vaccine, along with the immunostimulant AS02B. While disease free survival was increased in melanoma patients showing gene signature patterns correlative with pre-metastasis, the increase was not significant compared to patients treated with adjuvant alone (HR 0.42 95% CI 0.17 to 1.03 $P = 0.06$). Moreover, no change was seen in patients without pre-metastatic signatures (HR 1.17 95% CI 0.59 to 2.31 $P = 0.65$) [10]. In a separate study, investigating stages III and IV melanoma ($N = 42$) breast ($N = 1$), and bladder cancer ($N = 1$), NY-ESO-1 peptide vaccine was administered with dose [14], an antigen-specific T-cell response was detected in 6 treated patients [15]. However, a follow up study determined little difference in the decrease of T-regulatory T-cells after NY-ESO-1 vaccination [15]. Therefore, despite positive results seen in early clinical trials, vaccines against overexpressed self-antigens have not been successful in later stage clinical trials.

The other method to identify antigens for therapeutic cancer vaccines involves targeting tumor-specific mutations. In both clinical and pre-clinical models, these neoantigens have been shown to stimulate an immune response [16,17]. Furthermore, these mutations are less likely to develop immunologic tolerance because they have not been previously seen by the immune system [18]. In a clinical trial with 25 melanoma patients, survivability as significantly correlated with the presence of neoantigens ($P < 0.001$ by log rank test) [17]. Furthermore, 3 of 5 patients analyzed demonstrated substantial amounts of interferon-gamma expression by CD107a+ immune cells, with one responder showing an immunogenic difference unique to the antigen mutation [17]. A therapeutic vaccine Targovax, targeting mutations of the KRAS gene in pancreatic cancer, is in Phase II clinical trials [19]. In another clinical trial, vaccination of resected pancreatic cancer patients with K-ras 21-mer peptides, containing a mutation at codon 12, demonstrated limited toxicity to the vaccine. However, the vaccine also demonstrated only a slight immune response when tested by delayed-type hypersensitivity; with only 1 in 24 patients showing an immune response and a median overall survival 20.3 months (95% CI, 11.6–45.3 months) [20]. Unfortunately, unique mutations are often specific to each tumor and therefore typically have to be developed specifically for each patient, making them impractical for a mass manufactured vaccine. Furthermore, the mutated peptides are frequently immunoedited often causing vaccines of these mutations to not be as useful when the clones containing the mutation are selected against by the tumor [21].

Overall, vaccines that would allow for one epitope to induce a "cascade" of other tumor associated antigens to be recognized (known as epitope spreading), elicit a Th1 instead of a Th2 or immunosuppressive response, and guard against T-cell anergy should increase the efficacy of rationally designed epitopes. In the following sections, we will evaluate current vaccine development techniques, propose a mechanism for development of rationally designed epitopes, as well as explain caveats in specific approaches for identifying antigens and designing vaccines.

3. Designing vaccines for effective cancer therapy

3.1. In silico *prediction of peptide binding affinity as an indicator of cancer vaccine efficacy*

Immunoinformatics allows for identification of peptides with the highest affinity to MHC complexes (pMHC), an event potentially necessary to induce T-cell activation [22–24]. Likewise, immunoinformatics may identify promiscuous epitopes by searching for commonalities in high binding sequences between polymorphic MHC alleles which allows epitope spreading [25–28]. These tools used in peptide based vaccine development include two types of *in silico* predictive methods to predict pMHC binding; sequence-based and structure based-methods.

3.1.1. *Sequence based predictions*

In sequence based predictions, epitope segments are analyzed using computer algorithm models that predict pMHC binding strengths within a given peptide. The prediction algorithms that are most popular utilize computer models that are built on data-driven comparisons that rely on the differences of the peptide-binding segments (motifs) among sequences. Altogether, these algorithms are known as sequence based algorithms [23,27], but can be further divided into subclasses based on the approach of the algorithm or the test data that is used. These subclasses include motif alignment/positioning of specific domains [23,29], quantitated matrix-based approaches [27], and machine learning-based algorithms [27].

Algorithms that specifically utilize motif alignment and positioning of specific domains are the most established methods used to predict the binding of epitopes to MHC receptors [23]. For example, one of the most used prediction tools SYFPEITHI employs this search mechanism. For these algorithms, the peptide amino acid sequence is queried for the presence of specific amino acid combinations that have already been physically demonstrated to be high-affinity binders to specific MHC alleles (motifs). Matrix-based positioning predictive tools (Quantitative matrix, QM) are similar to motif alignment because they use the peptide sequence but involve evaluating different sequence frames from a protein, developing matrix values based on the amino acid and their corresponding position, and then using the known physical binding data for each amino acid sequence to quantitatively predict the ability of each frame to bind to the MHC binding cleft (Tables 1 and 2)

While alignment-motif and QM based epitope predictions developed early models for calculating peptide/MHC binding, correlations between actual peptide/MHC affinity and the calculated values was low. This low prediction is thought to be due to lack of the ability to take into account competition with neighboring amino acids for the binding pockets of the MHC binding cleft, which is also known as a position weight matrix [30,31]. Machine learning algorithms address this by predicting the neighboring amino acid competition by affinity rankings. Predictions based on machine-learning techniques further fall into further subclasses: ant colony search strategy [32], artificial neural networks (ANNs), support vector machines (SVMs), and hidden markov models (HMMs) [23]. The HMM system is assumed to be a part of a markov process, a test process that occurs without knowing previous iteration results and uses a set of hidden variables being randomly compared to known variables. To develop a HMM for epitope prediction, a pool of peptides that had a confirmed ability to cause T-cell proliferation was used (MHCPEP) [33]. Using this pool, randomized multiple alignments were developed that are expected, but not known, to have a high binding to the binding cleft in MHC [34,35]. An algorithm for the binding is created by testing these random sequences against known sequences.

As their name implies, ant colony strategies (ACS, also known as ant colony optimization) are built upon structures that resemble how an ant interacts in its colony. For instance, an ant is looking for food it first randomly searches, but leaves behind a chemical trail that other ants follow. However, this trail evaporates over time, so the only trails that are left are the ones were many ants follow, presumably the most optimal trail to the food [32]. Like HMM, ACS uses multiple alignments of tested proteins to attempt to find the

Table 1
MHC class I prediction tools.

Database	Method/algorithm	AA length	URL
ProprediI	QM	9	http://www.imtech.res.in/raghava/propred1/
EpiMatrix	QM(Sturniolo)	9–10	http://www.epivax.com/epimatrix/
NetChop3.1	ANN	9	http://www.cbs.dtu.dk/services/NetChop/
KISS	SVM	9	http://cbio.ensmp.fr/kiss/
CTLPred	QM,ANN, SVM	9	http://www.imtech.res.in/raghava/ctlpred/
NETCTL	ANN	9	http://www.cbs.dtu.dk/services/NetCTL/
NETCTLpan	ANN	8–11	http://www.cbs.dtu.dk/services/NetCTLpan/
EpiJen	QM	9	http://www.ddg-pharmfac.net/epijen/EpiJen/EpiJen.htm
AntiJen	QM	9–10	http://www.ddg-pharmfac.net/antijen/AntiJen/antijenhomepage.htm
SVMHC	SVM	9–14	https://abi.inf.uni-tuebingen.de/Services/SVMHC/index_html
BIMAS	QM	8–10	http://www-bimas.cit.nih.gov/molbio/hla_bind/
EpiPred	BIMAS, SMVHC, SYFPETHI	8–11	http://etk.informatik.uni-tuebingen.de/
SYFPEITHI	Motif	8–11	http://www.syfpeithi.de/
PREDEP	QM	9	http://margalit.huji.ac.il/Teppred/mhc-bind/
BIMAS	QM	8–10	http://www-bimas.cit.nih.gov/molbio/hla_bind/
IEDB	netMHCPan, ANN, SMMPMBEC, SMM, pickpocket, net MHCCons	8–14	http://tools.immuneepitope.org/main/tcell/
ComPred	ANN/QM	9	http://www.imtech.res.in/raghava/nhlapred/comp.html
POPI 2.0	SVM	9+	http://iclab.life.nctu.edu.tw/POPI/
EPIMHC	Motif	6+	http://bio.dfci.harvard.edu/MIF/epimhc/epimhc.html
MOTIF_SCAN	Motif	9–11	http://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan?sample_input=1
ANNPRED	ANN	9	http://www.imtech.res.in/raghava/nhlapred/neural.html
SVRMHC	SVM (regression)	9+	http://svrmhc.biolead.org/
SVMHC	SVM	9–10	https://abi.inf.uni-tuebingen.de/Services/SVMHC/index_html
Rankpep	PSSM	8–11	http://imed.med.ucm.es/Tools/rankpep.html

most optimal binding sequence. However, in each iteration of the analysis pMHC binding will influence the categorization of the next peptide analyzed so that the process goes from random to the most optimized peptide for binding to the MHC complex. Using this feature the goal of ACS is to provide the optimal binding core within a peptide, which is suggested to be necessary for peptide MHC binding [32]. Likewise, while HMM traditionally uses MHCPEP as starting data to train its algorithm, ACS utilizes SYFPEITHI as its testing database [32,34]. When evaluating peptide MHC binding, this test peptide database from SYFPEITHI was used and compared to computations from thermodynamic models of changes in Gibbs free energy and found to be comparable [32].

Similar to the mimicking of ant behavior in ant colony algorithms, neural networks mimic the interconnection of neurons in the central nervous system. In the case of pMHC binding predictions, neural nets take up independent binding data (also known as "nodes"), from peptide segment combinations of a target protein using information about the chemical interactions of amino acid residues. These nodes are then tested against "training sets" for patterns among all of the combinations within the network that are constantly being tested, reformed, and scored in an area of the nodes known as the hidden layers [31,36,37]. These patterns are what make up the algorithms used for pattern recognition among the unknown peptides. While ANNs are great at detecting false positives/negatives, as well as showing the binders are have good binding affinity, they are innately less sensitive in distinguishing the border between true low affinity peptides and the false negatives [37]. This flaw could be the product of using training sets with too many high-affinity binders [38].

Support Vector Machines (SVMs) are machine learning algorithms that are able to analyze data sets that have multiple variables by a separation criterion called a hyperplane. During SVM analysis, multiple hyperplanes are analyzed for most optimal condition; a plane that, when populations are presented in three-dimension space, separates the groups by the greatest contrast [31,39]. In terms of prediction for epitopes for MHC receptors, the algorithm is trained by using test data can be separated in two classes: those peptides that stimulate an immune response and those that do not [40]. After multiple iterative rounds of optimization, the peptide segments are compared to similarity of the clustered "stimulator" and "non-stimulator" groups by scrutiny from the optimally developed hyperplane. In such a situation, SVM predictions demonstrate much higher sensitivity than ANN algorithms with only slightly less predictive value [40].

These machine learning techniques have been used in combination for analysis and prediction of multiple cancer vaccines. For

Table 2
MHC class II prediction tools.

Database	Method/algorithm	AA length	URL
NetMHCIIPan-3.0	Hobohm1	9–19	http://www.cbs.dtu.dk/services/NetMHCIIpan/
TEPITOPEpan	PSSM	9–25	http://www.biokdd.fudan.edu.cn/Service/TEPITOPEpan/index.html
PREDIVAC	SDR	9	http://predivac.biosci.uq.edu.au/
Propred	QM (Sturniolo)	9	http://www.imtech.res.in/raghava/propred/index.html
EpiMatrix	QM (Sturniolo)	9–10	http://www.epivax.com/epimatrix/
EpiDock	DS-QM	9	http://epidock.ddg-pharmfac.net/EpiDockPage.aspx
EpiTOP	QSAR	9	http://www.pharmfac.net/EpiTOP/
SVMHC	SVM/TEPITOPE	9–15	https://abi.inf.uni-tuebingen.de/Services/SVMHC/index_html
SYFPEITHI	Motif	15	http://www.syfpeithi.de/
EpiPredict	QM	9+	http://www.epipredict.de/index.html
Rankpep	PSSM	9	http://imed.med.ucm.es/Tools/rankpep.html
IEDB	NetMHCIIpan, Sturniolo, NN-align, SMM-align	15	http://tools.immuneepitope.org/mhcii/
POPI 2.0	SVM	9+	http://iclab.life.nctu.edu.tw/POPI/
HLADR4Pred	ANN/SVM	9	http://www.imtech.res.in/raghava/hladr4pred/
EPIMHC	Motif	9+	http://bio.dfci.harvard.edu/MIF/epimhc/epimhc.html

instance, ImMucin, a MUC1 based cancer vaccine, was predicted to be efficacious based on classification of pMHC affinity binding predictions formulated through a combination of a quantitative matrix-based tool (BIMAS) and ANN, SMM inclusive scores (IEDB) [41]. The authors found that vaccination with MUC1-SP-L showed increased proliferation of PBMC from 13 breast cancer patients as well as increased effector: target cytotoxicity via ^{35}S methionine release assay ($P < 0.01$). Furthermore, vaccinated T-cell lines showed increased responses to similar MUC1 epitopes, demonstrating cross-priming. When tested *in vivo* with DA-3TM tumor bearing mice, there was an increased survival in the vaccinated mouse cohort compared to the cohort treated with GMCSF adjuvant alone and significantly increased when compared to another MUC1 peptide vaccine plus adjuvant ($P < 0.01$) [41].

Newer peptide predictive algorithms are an amalgamation of multiple sequence alignments based on prediction with structural or physical data (electrochemical and thermodynamic data) known as specific-determining residue (SDR) based, or Quantitative Structure Activity Relationship (QSAR) regression, algorithms [27,42]. These hybrid algorithms of sequence based calculations and structure based epitope calculations, together known as pan-specific algorithms, make generalized binding forecasts over multiple MHC alleles with little to no pre-existing binding data for the epitope in question [43]. Historically, MHCII prediction has been hindered by the promiscuous nature of the MHCII binding with misalignment of the peptide due to the open binding cleft [27] as well as lack of biochemically- or structurally-based data that can backup predictions [44,45]. However, these pan-specific algorithms will take into account the binding promiscuity and polymorphic alleles of MHC II, allowing for better prediction rates [43]. As an alternative mechanism, one can find promiscuity values manually by ranking different peptides affinities based on a database that checks different alleles, where the epitope segments with the highest affinity across the different alleles is the most promiscuous and, therefore, has the highest potential of epitope spreading [25,26].

While sequence-based prediction tools are the most widely used method of *in silico* epitope prediction, they have been criticized as being outdated and having high rates of false positives and negatives [23]. Moreover, these algorithms do not take into account other activities that may be important for immunogenicity, such as pMHC complex stability, pMHC:T-cell receptor binding affinity [46,47], or even the role that epitope tertiary structure plays in T-cell activation [48]. Along these same lines, current predictive algorithms assume that higher avidity pMHC always correlates with greater immunogenicity, however, it has been demonstrated that extremely strong and weak affinities of peptides to MHC receptors can cause lack of activation, apoptosis, or tolerization [49–51]. The development of peptides that favor epitope spreading may be facilitated by utilizing systems based, wide-spread biological screening methods, known as "-omics" methods [52].

In epitope design, it is important to understand how the peptide will be presented on MHCI and MHCII. When dealing with MHCI or MHCII sequence-based predictions, certain additional considerations must be taken into account in regards to the biological process of antigen presentation as well as how the antigen is presented on the MHC. With MHCI epitope design, the pathway that antigens must undertake to be presented on an MHCI receptor has to be considered. This pathway involves processing peptides via proteasomal degradation, trafficking to the endoplasmic reticulum via TAP, endosomal processing [31], and affinity binding of a peptide to the MHCI complex. MHCII, on the other hand, do not undergo proteasomal processing via TAP. Moreover, MHCII's open binding pocket allows for peptides of 13–17 amino acids to be presented whereas MHCI peptide lengths are limited to 8–11 amino acids, with the N- and C-termini directly interacting with the MHCI molecules [31]. With these restrictions, several predictive tools have been made to look at these processes and predict the peptides ability to be presented as an antigen (Tables 1 and 2).

For MHCI tools, CTLPred uses a tool that allows prediction with an algorithm (QM, ANN, SVM, consensus, and combined) [31], NETCTL/NETCTLpan/NETChop are all available on the IEDB website [31]. NETCTL/NETCTLpan/NETChop is a web-based tool suite that combines the predictive mechanisms of pMHC binding affinity (NETCTL), predicting antigen presentation over a range of polymorphic MHCs (NETCTLpan) as well as predicting the presence after proteasomal cleavage [53] (Table 1). It should be noted that, because MHC class I bound antigens directly activate cytotoxic T cells, the antigens targeted should predominantly be expressed in cells that will need to be eliminated [54,55].

MHC Class II epitope presentation involves endosomal processing and direct presentation of the antigen to the MHCII in the late endosome but the complexity of antigen prediction in MHCII complexes comes from the open cleft that is formed in the MHCII binding grove allowing for epitope presentation of multiple lengths and orientations. Moreover, as listed in Table 2, certain predictive tools are able to handle the complexities found in MHCII epitope processing; such as the previously mentioned "pan-specific" algorithms.

3.1.2. Structure based predictions

Structural prediction tools use a combination of sequence information, binding data, homologous structural information (*i.e.* crystallographic data of known pMHC complexes), and computational methods for prediction of pMHC complexes [23]. These techniques also include simulations of ligand docking and screening of combinatorial peptide libraries which are common in computer-aided drug design [23]. Using an integration of these data, software-mediated simulations take all, or part, of a desired target protein and assay different 3D orientations of these peptide segments in the binding cleft of the MHC testing for ideal thermodynamic, and electrostatic complementarity.

In addition to docking based prediction approaches, protein threading (also known as non-homologous protein recognition) algorithms are used find areas that contain peptides that bind to MHC [23]. These threading predictions are remodeled over structural data backbones made from existing crystallographic structural data along with measurements of changes of thermodynamic favorability [23,56]. Using these data, threading predictions are able to scrutinize molecular dynamics to find the most thermodynamically favorable solution of the putative epitope inside a targeting MHC binding pocket [23]. In addition, structural and molecular modeling may be simplified by 3D modeling software, such as Discovery studio, Swiss PDB Viewer, and Sybyl along with thermodynamic program plug-ins to adjust peptides for the most favorable intermolecular energy [26]. Unfortunately, most of the modeling software requires the availability of Protein Data Bank (PDB) data, which may not be available for the epitope segment that is being queried.

While structurally-based epitope predictions have mostly been utilized to find epitopes of viral-like particles, interest in using this tool in the development of cancer vaccines is increasing. For example, the cancer testis antigens, MAGE-A6 and MAGE-A12, were analyzed for epitopes via structurally-based prediction tools. Structurally-based software packages, MODELLER and AutoDock, combined with the crystal structure obtained from the PDB, were compared to sequence-based prediction tool; BIMAS. From this analysis, the authors stimulated PBMC from 6 melanoma patients and found 3/11 peptides tested significantly increased interferon-gamma levels when compared to unstimulated controls (greater than 33% of all patients tested, $P < 0.05$) [57,58].

Structurally-based tools are known for their increased accuracy of predictive value. However, structurally-based algorithms

also have disadvantages including the need for high computational knowledge and a need for high amounts of computational power and time [23]. Furthermore, there is still a paucity of experimentally-determined crystallographic data that can be used as foundations for modeling predictions [23,31]. This may suggest that pan-specific based predictions, or machine-learning based methods, are more accurate and practical for prediction of peptide immunogenicity.

3.1.3. Using multiple algorithms to identify epitopes

Databases such as the Immune Epitope Database (IEDB) analysis website combine multiple algorithms to give a "composite score" known as a consensus score with the ability of looking at different peptide characteristics at the same time [29]. For instance, IEDB's "consensus method" score for MHC class II is made of up an amalgamation of the machine learning algorithms: support vector machines (SMM-Align), artificial neural networks (NN-align), as well as matrix-based approaches (Sturniolo/TEPITOPE). Using the consensus method, many researchers have developed peptide segments for vaccination. For example, peptide segments were developed for the tumor-associated antigen, kinesin family member 20A (KIF20A), using the consensus method rankings as selection criteria. Using these peptide based vaccines, the authors demonstrated increase in interferon-gamma expressing Th1 cells and KIF20A presenting dendritic cells *in vitro* as well as showing increased interferon expressing PBMC in pre-immunized patients with malignant head and neck tumor compared to healthy donors [59]. Unfortunately, the consensus method is only slightly better at predicting epitope-MHC binding affinity than SMM_align alone at discerning false positives for MHCII allele DRB1101 [29]. This observation may be due to the fact that information used in many of the epitope database predictions is collected from literature, from different experimental systems, which can deteriorate the consistency expected from training sets used to make sequence based algorithms [29]. Combining data analysis may overestimate the predictive value of both MHCI and II epitopes namely due to small training datasets that do not evenly cover the antigen sequence [60,61]. While these studies have only focused on the algorithms found in IEDB's analytical tools, meta-analysis to compare combined predicted algorithms to blind empirical control samples are important to test the robustness of the consensus of the combined algorithms (see Tables 1 and 2).

Currently the most promising methods of sequence-based peptide development is a meta-analysis of ranking peptide segments using a combination of methods that take into account sequence based binding predictions as well as spatial and physiochemical considerations of homologous peptides [43]. This analysis can be done by ranking peptides in terms of binding affinity and promiscuity, comparing predictions to structural data and diverse sets of test groups that are both similar and dissimilar to the peptide segments in question, followed by peptide validation [62]. However, while much binding data is available, there is also a paucity of data for certain functional events, such as the ability to undergo proteasomal cleavage before being loaded on the MHCI binding cleft [31]. Therefore, in order to benefit from *in silico* prediction, peptide development relies heavily on experimental validation techniques in the proceeding stages of vaccine development.

3.2. Other considerations

3.2.1. Alternative strategies to assaying peptide-MHC complex formation and immunogenicity

While computer based predictive values of pMHC affinity offers an inexpensive method for determination of immunogenic epitopes, other methods (such as immunoproteomics) offer an alternative that is able to determine pMHC binding in specific situations. Immunoproteomics involves that isolation of peptides from pMHC complexes of the tumor, isolation of the peptide bound to the MHC complex followed by identification of the peptide structure by high performance liquid chromatography (HPLC) fractionation and mass spectrometry [55,63].

Likewise, *in vivo* discernment of the immunogenicity of selected epitopes is usually demonstrated *in vitro* first, using cytokine expressing assays (flow Cytometry, qRT-PCR, ELISA, and ELISPOT) or via target: effector tests (CD8+ Chromium release assay, flow-cytometry based target:effector tests, expression of Granzymes by CD8+ T-cells). These *in vitro* screens are followed by *in vivo* studies of tumor regression, *ex-vivo* analysis of immune cell infiltration (via flow Cytometry or microscopy) as well as *ex-vivo* examination of the presence of a Th1/Th2 effector response (ELISA, ELISPOT, intracellular cytokine flow) as well as T-cell activation (flow cytometry).

4. Conclusion

Current vaccine-based therapies for cancer have demonstrated immune responses in multiple tumor types. However, lack of controlled immune responses may limit the efficacy of cancer vaccines. Rationally-designed vaccine epitopes show promise to produce this controlled immune response, but there are still multiple limitations for designing cancer vaccine epitopes. As peptide libraries of structural data increase, and more insight is gained into the process of MHCI and MHCII antigen presentation, rational design of peptide-based vaccines may see an increase in sophistication which may lead to more effective cancer immunotherapy.

Conflict of interest

Sasha E. Stanton and Gregory A. Cherryholmes have no conflicts of interest. Mary L. Disis' conflicts of interest are as follows: Grant support: Celgene, VentiRx, Seattle Genetics, EMD Serono; Stock: VentiRx, Epithany.

References

[1] Fridman WH, Pagès F, Sautès-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. Nat Rev Cancer 2012;12:298–306, http://dx.doi.org/10.1038/nrc3245.
[2] Vonderheide RH. Prospects and challenges of building a cancer vaccine targeting telomerase. Biochimie 2008;90:173–80, http://dx.doi.org/10.1016/j.biochi.2007.07.005.
[3] Hardacre JM, Mulcahy MF, Small W, Talamonti MS, Obel JC, Rocha Lima CMS, et al. Addition of algenpantucel-L immunotherapy to standard of care (SOC) adjuvant therapy for pancreatic cancer. ASCO Annu Meet Proc 2012;30:4049.
[4] Hardacre JM, Mulcahy M, Small W, Talamonti M, Obel J, Krishnamurthi S, et al. Addition of algenpantucel-L immunotherapy to standard adjuvant therapy for pancreatic cancer: a phase 2 study. J Gastrointest Surg 2013;17:94–101, http://dx.doi.org/10.1007/s11605-012-2064-6.
[5] Ouellette MM, Wright WE, Shay JW. Targeting telomerase-expressing cancer cells. J Cell Mol Med 2011;15:1433–42, http://dx.doi.org/10.1111/j. 1582-4934.2011.01279.x.
[6] Carmichael MG, Benavides LC, Holmes JP, Gates JD, Mittendorf EA, Ponniah S, et al. Results of the first phase 1 clinical trial of the HER-2/neu peptide (GP2) vaccine in disease-free breast cancer patients. Cancer 2010;116:292–301, http://dx.doi.org/10.1002/cncr.24756.
[7] Lee J, Kim KM, Kang W, Ou SH. Innovative personalized medicine in gastric cancer: time to move forward. Clin Genet 2014;86:37–43, http://dx.doi.org/10.1111/cge.12408.
[8] Owens MA, Horten BC, Da Silva MM. HER2 amplification ratios by fluorescence in situ hybridization and correlation with immunohistochemistry in a cohort of 6556 breast cancer tissues. Clin Breast Cancer 2004;5:63–9.
[9] Fratta E, Coral S, Covre A, Parisi G, Colizzi F, Danielli R, et al. The biology of cancer testis antigens: putative function, regulation and therapeutic potential. Mol Oncol 2011;5:164–82.
[10] Ulloa-Montoya F, Louahed J, Dizier B, Gruselle O, Spiessens B, Lehmann FF, et al. Predictive gene signature in MAGE-A3 antigen-specific cancer immunotherapy. J Clin Oncol 2013;31:2388–95, http://dx.doi.org/10.1200/JCO.2012.44.3762.
[11] Vansteenkiste J, Zielinski M, Linder A, Dahabre J, Esteban E, Malinowski W, et al. Final results of a multi-center, double-blind, randomized, placebo-controlled phase II study to assess the efficacy of MAGE-A3 immunotherapeutic as

adjuvant therapy in stage IB/II non-small cell lung cancer (NSCLC). ASCO Annu Meet Proc 2007;25:7554.
[12] de Carvalho F, Vettore AL, Inaoka RJ, Karia B, Andrade VC, Gnjatic S, et al. Evaluation of LAGE-1 and NY-ESO-1 expression in multiple myeloma patients to explore possible benefits of their homology for immunotherapy. Cancer Immun: J Acad Cancer Immunol 2011;11:11.
[13] Vansteenkiste J, Zielinski M, Linder A, Dahabreh J, Gonzalez EE, Malinowski W, et al. Adjuvant MAGE-A3 immunotherapy in resected non-small-cell lung cancer: phase II randomized study results. J Clin Oncol 2013;31:2396–403, http://dx.doi.org/10.1200/JCO.2012.43.7103.
[14] Davis ID, Chen W, Jackson H, Parente P, Shackleton M, Hopkins W, et al. Recombinant NY-ESO-1 protein with ISCOMATRIX adjuvant induces broad integrated antibody and CD4+ and CD8+ T cell responses in humans. Proc Natl Acad Sci USA 2004;101:10697–702, http://dx.doi.org/10.1073/pnas.0403572101.
[15] Klein O, Davis ID, McArthur GA, Chen L, Haydon A, Parente P, et al. Low-dose cyclophosphamide enhances antigen-specific CD4+ T cell responses to NY-ESO-1/ISCOMATRIX™ vaccine in patients with advanced melanoma. Cancer Immunol Immunother 2015:1–12, http://dx.doi.org/10.1007/s00262-015-1656-x.
[16] DuPage M, Cheung AF, Mazumdar C, Winslow MM, Bronson R, Schmidt LM, et al. Endogenous T cell responses to antigens expressed in lung adenocarcinomas delay malignant tumor progression. Cancer Cell 2011;19:72–85, http://dx.doi.org/10.1016/j.ccr.2010.11.011.
[17] Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med 2014;371:2189–99, http://dx.doi.org/10.1056/NEJMoa1406498.
[18] Khalili JS, Hanson RW, Szallasi Z. In silico prediction of tumor antigens derived from functional missense mutations of the cancer gene census. OncoImmunology 2012;1:1281, http://dx.doi.org/10.4161/onci.21511.
[19] Melero I, Gaudernack G, Gerritsen W, Huber C, Parmiani G, Scholl S, et al. Therapeutic vaccines for cancer: an overview of clinical trials. Nat Rev Clin Oncol 2014;11:509–24, http://dx.doi.org/10.1038/nrclinonc.2014.111.
[20] Abou-Alfa GK, Chapman PB, Feilchenfeldt J, Brennan MF, Capanu M, Gansukh B, et al. Targeting mutated K-ras in pancreatic adenocarcinoma using an adjuvant vaccine. Am J Clin Oncol 2011;34:321–5, http://dx.doi.org/10.1097/COC.0b013e3181e84b1f.
[21] Khodadoust MS, Alizadeh AA. Tumor antigen discovery through translation of the cancer genome. Immunol Res 2014;58:292–9, http://dx.doi.org/10.1007/s12026-014-8505-4.
[22] Kumar V, Bhardwaj V, Soares L, Alexander J, Sette A, Sercarz E. Major histocompatibility complex binding affinity of an antigenic determinant is crucial for the differential secretion of interleukin 4/5 or interferon gamma by T cells. Proc Natl Acad Sci 1995;92:9510–4.
[23] Patronov A, Doytchinova I. T-cell epitope vaccine design by immunoinformatics. Open Biol 2013;3:120139, http://dx.doi.org/10.1098/rsob.120139.
[24] Sercarz EE, Lehmann PV, Ametani A, Benichou G, Miller A, Moudgil K. Dominance and crypticity of T cell antigenic determinants. Annu Rev Immunol 1993;11:729–66.
[25] Hu Y, Petroni GR, Olson WC, Czarkowski A, Smolkin ME, Grosh WW, et al. Immunologic hierarchy, class II MHC promiscuity, and epitope spreading of a melanoma helper peptide vaccine. Cancer Immunol Immunother 2014:1–8, http://dx.doi.org/10.1007/s00262-014-1551-x.
[26] Mishra S, Sinha S. Immunoinformatics, molecular modeling, and cancer vaccines. Meth Mol Biol 2014;1184:513–21.
[27] Oyarzún P, Ellis JJ, Bodén M, Kobe B. PREDIVAC. CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II DR protein diversity. BMC Bioinf 2013;14:52, http://dx.doi.org/10.1186/1471-2105-14-52.
[28] zur Wiesch JS, Lauer GM, Day CL, Kim AY, Ouchi K, Duncan JE, et al. Broad repertoire of the CD4+ Th cell response in spontaneously controlled hepatitis C virus infection includes dominant and highly promiscuous epitopes. J Immunol 2005;175:3603–13, http://dx.doi.org/10.4049/jimmunol.175.6.3603.
[29] Wang P, Sidney J, Dow C, Mothe B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol 2008;4:e1000048, http://dx.doi.org/10.1371/journal.pcbi.1000048.
[30] 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;12:1007–17, http://dx.doi.org/10.1110/ps.0239403.
[31] Desai DV, Kulkarni-Kale U. T-cell epitope prediction methods: an overview. Meth Mol Biol 2014;1184:333–64.
[32] Karpenko O, Shi J, Dai Y. Prediction of MHC class II binders using the ant colony search strategy. Artif Intell Med 2005;35:147–56.
[33] Brusic V, Rudy G, Harrison LC. MHCPEP—a database of MHC-binding peptides: update 1997. Nucleic Acids Res 1998;26:368–71, http://dx.doi.org/10.1093/nar/26.1.368.
[34] Mamitsuka H. Predicting peptides that bind to MHC molecules using supervised learning of hidden Markov models. Proteins Struct Funct Genet 1998;33:460–74.
[35] 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;94:264–70.
[36] Honeyman MC, Brusic V, Stone NL, Harrison LC. Neural network-based prediction of candidate T-cell epitopes. Nat Biotechnol 1998;16:966–9, http://dx.doi.org/10.1038/nbt1098-966.
[37] Gulukota K, Sidney J, Sette A, DeLisi C. Two complementary methods for predicting peptides binding major histocompatibility complex molecules. J Mol Biol 1997;267:1258–67.
[38] Jørgensen KW, Rasmussen M, Buus S, Nielsen M. NetMHCstab—predicting stability of peptide–MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. Immunology 2014;141:18–26, http://dx.doi.org/10.1111/imm.12160.
[39] Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Cambridge: Cambridge University Press; 2000.
[40] Zhao Y, Pinilla C, Valmori D, Martin R, Simon R. Application of support vector machines for T-cell epitopes prediction. Bioinformatics 2003;19:1978–84, http://dx.doi.org/10.1093/bioinformatics/btg255.
[41] Kovjazin R, Volovitz I, Kundel Y, Rosenbaum E, Medalia G, Horn G, et al. ImMucin: a novel therapeutic vaccine with promiscuous MHC binding for the treatment of MUC1-expressing tumors. Vaccine 2011;29:4676–86, http://dx.doi.org/10.1016/j.vaccine.2011.04.103.
[42] Doytchinova IA, Flower DR. QSAR and the prediction of T-cell epitopes. Curr Proteomics 2008;5:73–95. ⟨http://dx.doi.org/10.2174/157016408784911945⟩.
[43] Zhang H, Wang P, Papangelopoulos N, Xu Y, Sette A, Bourne PE, et al. Limitations of Ab initio predictions of peptide binding to MHC class II molecules. PLoS ONE 2010;5:e9272, http://dx.doi.org/10.1371/journal.pone.0009272.
[44] Khan JM, Kumar G, Ranganathan S. RESEARCH open access In silico prediction of immunogenic T cell epitopes for HLA-DQ8. Immunome Res 2012;8.
[45] Khan JM, Tong JC, Ranganathan S. Structural immunoinformatics: understanding MHC–peptide–TR binding. In: Bioinformatics for immunomics. New York, NY: Springer; 2010. p. 77–93.
[46] Holler PD, Kranz DM. Quantitative analysis of the contribution of TCR/pepMHC affinity and CD8 to T cell activation. Immunity 2003;18:255–64.
[47] Corse E, Gottschalk RA, Allison JP. Strength of TCR–peptide/MHC interactions and in vivo T cell responses. J Immunol 2011;186:5039–45, http://dx.doi.org/10.4049/jimmunol.1003650.
[48] Ma C, Whiteley PE, Cameron PM, Freed DC, Pressey A, Chen S-L, et al. Role of APC in the selection of immunodominant T cell epitopes. J Immunol 1999;163:6413–23.
[49] Snyder JT, Alexander-Miller MA, Berzofskyl JA, Belyakov IM. Molecular mechanisms and biological significance of CTL avidity. Curr HIV Res 2003;1:287–94, http://dx.doi.org/10.2174/1570162033485230.
[50] Scardino A, Gross D-A, Alves P, Schultze JL, Graff-Dubois S, Faure O, et al. HER-2/neu and hTERT cryptic epitopes as novel targets for broad spectrum tumor immunotherapy. J Immunol 2002;168:5900–6, http://dx.doi.org/10.4049/jimmunol.168.11.5900.
[51] Gervois N, Guilloux Y, Diez E, Jotereau F. Suboptimal activation of melanoma infiltrating lymphocytes (TIL) due to low avidity of TCR/MHC-tumor peptide interactions. J Exp Med 1996;183:2403–7, http://dx.doi.org/10.1084/jem.183.5.2403.
[52] Disis ML. Immunologic biomarkers as correlates of clinical response to cancer immunotherapy. Cancer Immunol Immunother 2011;60:433–42, http://dx.doi.org/10.1007/s00262-010-0960-8.
[53] Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC bioinf 2007;8:424, http://dx.doi.org/10.1186/1471-2105-8-424.
[54] Fortier M-H, Caron É, Hardy M-P, Voisin G, Lemieux S, Perreault C, et al. The MHC class I peptide repertoire is molded by the transcriptome. J Exp Med 2008;205:595–610, http://dx.doi.org/10.1084/jem.20071985.
[55] Comber JD, Philip R. MHC class I antigen presentation and implications for developing a new generation of therapeutic vaccines. Ther Adv Vaccines 2014;2:77–89, http://dx.doi.org/10.1177/2051013614525375.
[56] Akutsu T, Sim KL. Protein threading based on multiple protein structure alignment. Genome Inf 1999;10:23–9, http://dx.doi.org/10.11234/gi1990.10.23.
[57] Nakamura Y, Tai S, Oshita C, Iizuka A, Ashizawa T, Saito S, et al. Analysis of HLA-A24-restricted peptides of carcinoembryonic antigen using a novel structure-based peptide-HLA docking algorithm. Cancer Sci 2011;102:690–6, http://dx.doi.org/10.1111/j. 1349-7006.2011.01866.x.
[58] Akiyama Y, Komiyama M, Nakamura Y, Iizuka A, Oshita C, Kume A, et al. Identification of novel MAGE-A6-and MAGE-A12-derived HLA-A24-restricted cytotoxic T lymphocyte epitopes using an in silico peptide-docking assay. Cancer Immunol Immunother 2012;61:2311–9, http://dx.doi.org/10.1007/s00262-012-1298-1.
[59] Tomita Y, Yuno A, Tsukamoto H, Senju S, Kuroda Y, Hirayama M, et al. Identification of promiscuous KIF20A long peptides bearing both CD4+ and CD8+ T-cell epitopes: KIF20A-specific CD4+ T-cell immunity in patients with malignant tumor. Clin Cancer Res 2013;19:4508–20, http://dx.doi.org/10.1158/1078-0432.CCR-13-0197.
[60] Kim Y, Sidney J, Buus S, Sette A, Nielsen M, Peters B. Dataset size and composition impact the reliability of performance benchmarks

for peptide-MHC binding predictions. BMC Bioinf 2014;15:241, http://dx.doi.org/10.1186/1471-2105-15-241.

[61] Chaves FA, Lee AH, Nayak JL, Richards KA, Sant AJ. The utility and limitations of current Web-available algorithms to predict peptides recognized by CD4 T cells in response to pathogen infection. J Immunol 2012;188:4235–48, http://dx.doi.org/10.4049/jimmunol.1103640.

[62] Six A, Bellier B, Thomas-Vaslin V, Klatzmann D. Systems biology in vaccine design. Microb Biotechnol 2012;5:295–304, http://dx.doi.org/10.1111/j. 1751-7915.2011.00321.x.

[63] Purcell AW, Gorman JJ. Immunoproteomics mass spectrometry-based methods to study the targets of the immune response. Mol Cell Proteomics 2004;3:193–208, http://dx.doi.org/10.1074/mcp.R300013-MCP200.