Vaccine xxx (2015) xxx–xxx

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 xxx

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

Please cite this article in press as: Cherryholmes GA, et al. Current methods of epitope identification for cancer vaccine design. Vaccine (2015), http://dx.doi.org/10.1016/j.vaccine.2015.06.116