9-15-2015

Development and Validation of an Epitope Prediction Tool for Swine (PigMatrix) Based on the Pocket Profile Method

Andres H. Gutiérrez

William D. Martin

Chris Bailey-Kellogg

Frances Terry

Leonard Moise
University of Rhode Island, lmoise@uri.edu

See next page for additional authors

Development and Validation of an Epitope Prediction Tool for Swine (PigMatrix) Based on the Pocket Profile Method

Creative Commons License

Authors

Andres H. Gutiérrez, William D. Martin, Chris Bailey-Kellogg, Frances Terry, Leonard Moise, and Anne S. De Groot

Terms of Use

BMC Bioinformatics

CrossMark

Development and validation of an epitope prediction tool for swine (PigMatrix) based on the pocket profile method

Andres H. Gutiérrez[1], William D. Martin[2], Chris Bailey-Kellogg[3], Frances Terry[2], Leonard Moise[1,2] and Anne S. De Groot[1,2]*

Abstract

Background: T cell epitope prediction tools and associated vaccine design algorithms have accelerated the development of vaccines for humans. Predictive tools for swine and other food animals are not as well developed, primarily because the data required to develop the tools are lacking. Here, we overcome a lack of T cell epitope data to construct swine epitope predictors by systematically leveraging available human information. Applying the "pocket profile method", we use sequence and structural similarities in the binding pockets of human and swine major histocompatibility complex proteins to infer Swine Leukocyte Antigen (SLA) peptide binding preferences. We developed epitope-prediction matrices (PigMatrices), for three SLA class I alleles (SLA-1*0401, 2*0401 and 3*0401) and one class II allele (SLA-DRB1*0201), based on the binding preferences of the best-matched Human Leukocyte Antigen (HLA) pocket for each SLA pocket. The contact residues involved in the binding pockets were defined for class I based on crystal structures of either SLA (SLA-specific contacts, Ssc) or HLA supertype alleles (HLA contacts, Hc); for class II, only Hc was possible. Different substitution matrices were evaluated (PAM and BLOSUM) for scoring pocket similarity and identifying the best human match. The accuracy of the PigMatrices was compared to available online swine epitope prediction tools such as PickPocket and NetMHCpan.

Results: PigMatrices that used Ssc to define the pocket sequences and PAM30 to score pocket similarity demonstrated the best predictive performance and were able to accurately separate binders from random peptides. For SLA-1*0401 and 2*0401, PigMatrix achieved area under the receiver operating characteristic curves (AUC) of 0.78 and 0.73, respectively, which were equivalent or better than PickPocket (0.76 and 0.54) and NetMHCpan version 2.4 (0.41 and 0.51) and version 2.8 (0.72 and 0.71). In addition, we developed the first predictive SLA class II matrix, obtaining an AUC of 0.73 for existing SLA-DRB1*0201 epitopes. Notably, PigMatrix achieved this level of predictive power without training on SLA binding data.

Conclusion: Overall, the pocket profile method combined with binding preferences from HLA binding data shows significant promise for developing T cell epitope prediction tools for pigs. When combined with existing vaccine design algorithms, PigMatrix will be useful for developing genome-derived vaccines for a range of pig pathogens for which no effective vaccines currently exist (e.g. porcine reproductive and respiratory syndrome, influenza and porcine epidemic diarrhea).

Keywords: PigMatrix, EpiMatrix, Computational vaccinology, Epitope prediction, HLA, SLA, MHC, Class I, Class II, Porcine, PRRSV, Influenza, Genome-derived vaccine, T cell epitope

* Correspondence: dr.annie.degroot@gmail.com
[1]Institute for Immunology and Informatics, CMB/CELS, University of Rhode Island, Providence, RI 02903, USA
[2]EpiVax, Inc., Providence, RI 02860, USA
Full list of author information is available at the end of the article

Gutiérrez *et al. BMC Bioinformatics* (2015) 16:290

Page 2 of 13

Background

The interaction of Major Histocompatibility Complex (MHC) proteins with peptides derived from protein antigens plays a key role in the adaptive immune response mediated by T cells. The MHC:peptide complex presented on the surface of a cell is recognized by the T cell receptor (TCR), which activates the T cell and drives the immune response. There are two classes of MHC molecules: MHC class I presents peptides of intracellular origin to CD8+ T cells (cytotoxic T cells, or CTL) and MHC class II presents peptides of extracellular origin to CD4+ T cells (T-helper cells, or Th). Both classes of molecules have similar tertiary structure. Class I molecules have a transmembrane (α) chain noncovalently associated with β₂-microglobulin where the α₁ and α₂ domains form the peptide-binding groove; class II molecules have two transmembrane chains (α and β) where the α₁ and β₁ domains form the peptide-binding groove. The MHC class I binding groove is closed, which restricts the length of bound peptides to 8–10 residues; the MHC class II binding groove on the other hand, is open, and peptides can extend beyond the ends of the groove, allowing binding of longer and more flexible peptides of variable lengths (typically 13–25 amino acids) [1].

The tertiary structure of MHC molecules is relatively conserved, even across species. For example, crystallographic studies have shown similarity between Human Leukocyte Antigen (HLA; human MHC) and Swine Leukocyte Antigen (SLA; swine MHC) molecules [2]. The SLA-1*0401 class I allele has been crystallized in complex with peptides derived from 2009-pandemic H1N1 (pH1N1) swine-origin influenza A virus (S-OIV) and Ebola virus. A structural comparison revealed that the SLA class I molecule, SLA-1*0401 contains six pockets in its binding groove, similar to HLA class I molecules. The root-mean-squared deviation (RMSD) for all of the Cα atoms in SLA-1*0401 and HLA-A*1101, which has the highest identity with SLA-1*0401 (78 %), was <0.7 Å indicating a similar arrangement of their backbones. Furthermore, three out of 23 influenza SLA-1*0401 binders were identical to previously defined peptides presented by HLA-A*0101 [2]. For SLA class II, no crystal structures are available, but amino acid SLA-DR sequences are highly similar to their human counterparts. For example, the amino acid sequences of SLA-DRB1*0201 and HLA-DRB1*0101 are 79 % identical.

Due to the importance of peptide binding to MHC molecules in the immune response, human T cell epitope prediction tools have been developed based on a range of approaches and are widely used in vaccine development and experimental immunology [3]. The availability of a large and expanding database of validated MHC ligands has contributed to the development of more accurate algorithms. Epitope predictions using these tools reduce the time and effort required to identify T cell epitopes [4]. The number of epitope prediction tools is more limited for pigs due to the paucity of experimental data available. To overcome the lack of quantitative measurements of MHC interaction for a large number of HLA alleles, 'pan-specific' methods have been implemented for prediction of T cell epitopes that bind to MHC for which experimental data are limited or not available. Pan-specific methods use experimental binding data and amino acid sequences of multiple MHC alleles to infer binding preferences to uncharacterized MHC molecules. These methods have been used for development of prediction tools for MHC class I [5–7] and II alleles [8–11], but only NetMHCpan has been used for prediction of SLA class I-restricted peptides [12–16]. This method is based on artificial neural networks (ANN) trained using as input a pseudo-sequence composed of the polymorphic residues in the binding groove of a given MHC, a peptide sequence and the experimental affinity data. To our knowledge, there are no *in silico* tools that are available for SLA class II.

Sturniolo et al. first described a method for using existing data to develop new epitope predictors, the pocket profile method, in 1999 [17]. It has been used to develop pan-specific methods for predicting binding of peptides to HLA class I and II alleles [9, 18]. The approach depends on the identification of certain polymorphic regions within HLA molecules that are known to be the areas of contact between peptides and the binding groove of HLA [19–21]. Contact residues from the HLA molecule that bind the R group (side chain) of a specific amino acid within a linear peptide can be considered to form a pocket for that R group. Thus, each 'pocket' can also be described in terms of its amino acid binding preferences ('pocket profile'). The pocket profiles are nearly independent of the remaining binding groove. So this method assumes that two MHC alleles with identical pocket residues will have the same pocket profile. Therefore, given sufficient information about the contact residues of the set of pockets in the binding groove of an MHC and experimentally determined pocket profiles, it is possible to compose predictive matrices *in silico*. The method was originally applied to develop TEPITOPE, an algorithm for prediction of peptide ligands to 51 HLA class II alleles with known pocket residues [17] and then extended to any HLA-DR molecules with similar pockets (TEPITOPEpan) [9]. A similar method has also been used in the PickPocket algorithm for MHC class I prediction [18]. Whereas TEPITOPEpan uses pocket profiles from TEPITOPE, PickPocket generates binding preferences using position-specific scoring matrices (PSSMs) from binding data directly. Although no publications exist using these algorithms for SLA binding predictions, SLA alleles are available for use

Gutiérrez *et al. BMC Bioinformatics* (2015) 16:290

Page 3 of 13

in the PickPocket server (www.cbs.dtu.dk/services/Pick-Pocket/).

EpiMatrix is a matrix-based algorithm that uses the pocket profile method to predict potential HLA class I and II T cell epitopes. The first version of this algorithm was developed in 1996, and newer versions have been extensively validated *in vitro* in HLA binding assay and human T cell assays and in animal studies using HLA transgenic mouse models [22–26]. For common class II alleles, EpiMatrix appears to predict more accurately than many available epitope-mapping algorithms [27]. Comparative performance for EpiMatrix class I predictions has not been published; however, the tools have been successfully applied to identify class I-restricted T cell epitopes in human pathogens [28–30]. The pocket profile method was used to develop a matrix for a bovine MHC class I allele [31] and in the early 2000s, this method was also used to derive SLA class II prediction matrices from EpiMatrix, but this work was not published.

This paper describes the development and retrospective validation of predictive matrices to map T cell epitopes for SLA class I (SLA-1*0401, 2*0401, 3*0401) alleles and a class II (SLA-DRB1*0201) allele. "PigMatrix" matrices are built based on the pocket profile method using EpiMatrix pocket profiles for HLA epitope prediction. While these alleles represent a small subset of commonly expressed alleles in pigs [32–36], they were selected for their available peptide binding data [13–16, 37–39]. As before, we assumed that predictive matrices developed for HLA alleles should function as reasonable proxies for the prediction of ligands to SLA molecules with similar pocket profiles. Thus, we developed 'composite matrices' by selecting the most similar HLA pocket (best human match) for each SLA pocket, and built matrices composed of the corresponding HLA binding preferences (Fig. 1).

Two methods were used to define the pocket contact residues considering different scenarios of availability of SLA crystal structures. In the first scenario, SLA crystal structures were available, so pockets were defined from these structures. In the second, no SLA crystal structures were available; therefore, contact residues were selected based on crystal structures of HLA. We also tested different substitution matrices (PAM and BLOSUM) to score pocket similarity to define the best human match. PigMatrix was benchmarked against existing SLA prediction tools for class I alleles. Benchmarking against other SLA class II predictors was not possible as no other prediction algorithms are available. The results demonstrate the potential of this approach to develop matrices to make accurate predictions for both SLA class I and II alleles for which experimental binding data are limited or even non-existent.
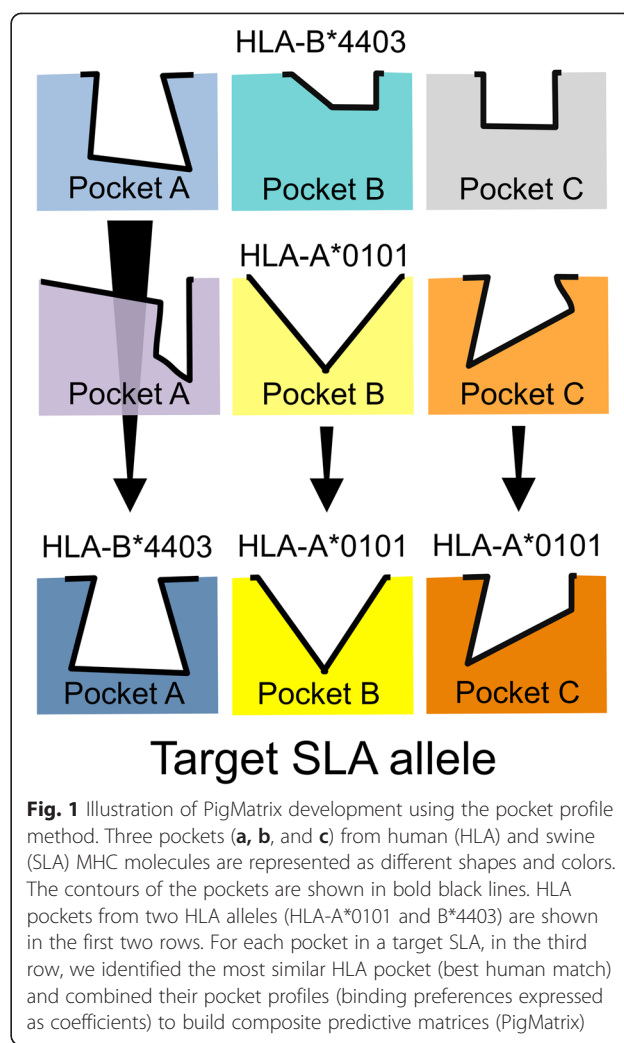

Fig. 1 Illustration of PigMatrix development using the pocket profile method. Three pockets (**a, b**, and **c**) from human (HLA) and swine (SLA) MHC molecules are represented as different shapes and colors. The contours of the pockets are shown in bold black lines. HLA pockets from two HLA alleles (HLA-A*0101 and B*4403) are shown in the first two rows. For each pocket in a target SLA, in the third row, we identified the most similar HLA pocket (best human match) and combined their pocket profiles (binding preferences expressed as coefficients) to build composite predictive matrices (PigMatrix)

Methods

Datasets

Unique 9-mer peptides with reported binding measurement to a specific SLA allele were compiled from the literature into two datasets: one comprising binders and the other, non-binders, for each of three class I (SLA-1*0401, 2*0401, 3*0401) alleles [2, 13–16]. The SLA-1*0401 dataset included 133 binders and 46 non-binders; 2*0401 included 24 binders and 46 non-binders; and 3*0401, 27 binders and 46 non-binders. Twenty-five (14 %) of the SLA-1*0401 peptides were reported by Zhang et al. [2]; the remaining peptides for 1*0401, 2*0401 and 3*0401 were published by Pedersen et al. in different publications [13–16]. For class II, a dataset was created with peptides specific to SLA-DRB1*0201 from the literature [37–39]. This dataset has 33 binders and 171 non-binders. Peptides with contradictory (both positive and negative) results were discarded (Additional file 1). Additionally, we generated a set of 100,000 unique 9-mer peptides from random sequence

Gutiérrez *et al. BMC Bioinformatics* (2015) 16:290

Page 4 of 13

proteins with the average amino acid frequencies of the proteins in the Swiss-Prot database for use as a control data set, as previously described [31]. The random proteins were generated using the RandSeq tool from ExPASy [40].
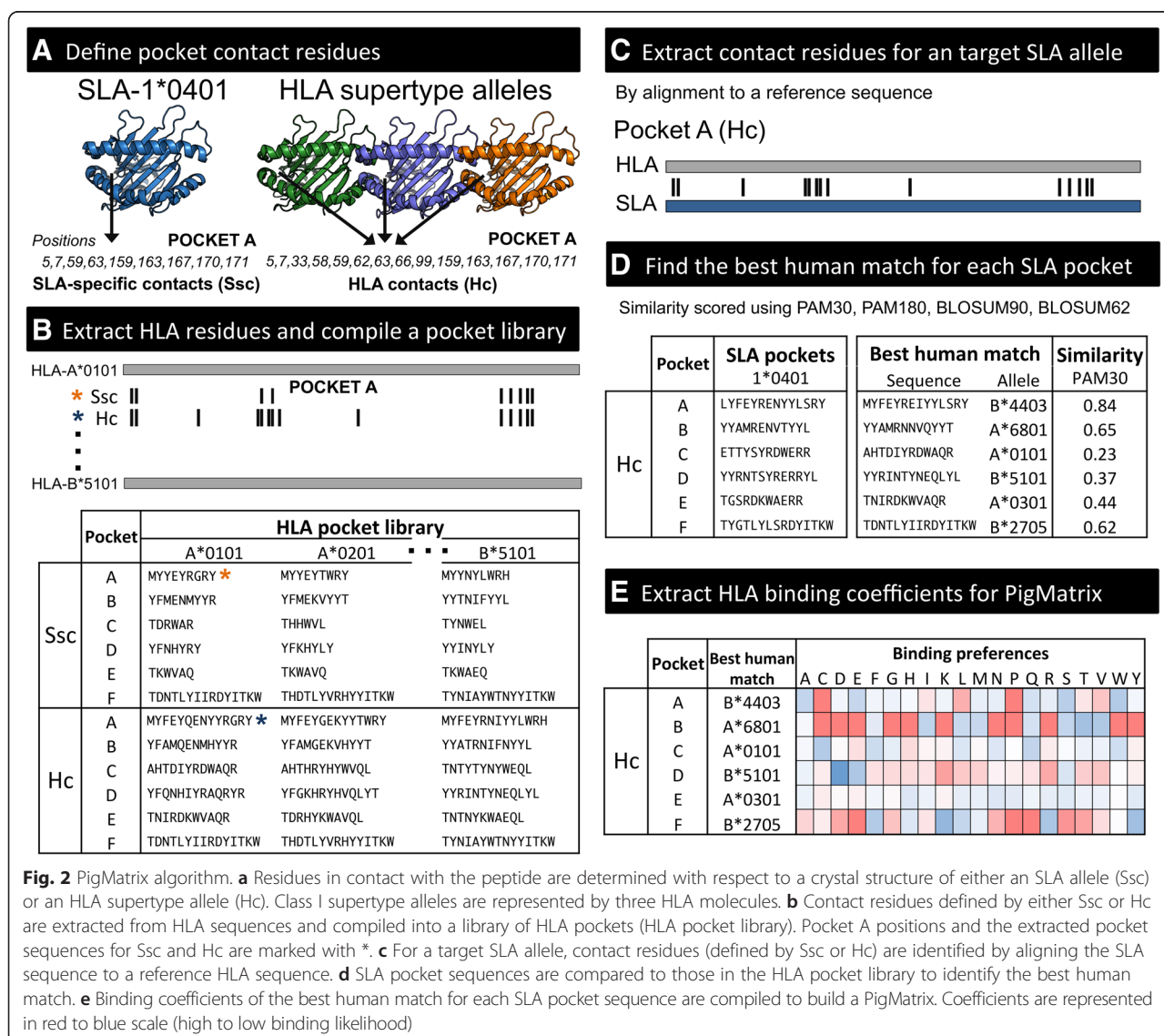
MHC sequences

Complete amino acid sequences from SLA protein sequences, along with HLA class I (HLA-A*0101, A*0201, A*0301, A*1101, A*2402, A*6801, B*0702, B*0801, B*2705, B*3501, B*4403, B*5101) and class II alleles (HLA-DRB1*0101, 0301, 0401, 0701, 0801, 1101, 1301, 1501), were obtained from the IPD-MHC Database (www.ebi.ac.uk/ipd/). It is important to clarify that the HLA alleles for which binding preferences are available in EpiMatrix are families of alleles that share pocket preferences, rather than individual alleles. The alleles

represent 12 class I supertypes [41] and eight class II supertypes [42].

Binding pocket residues

Six pockets (A-F) and five pockets (A-E) were considered for class I and II, respectively. Pockets for peptide positions 4, 5 and 8 for class I and 2, 3, 5, and 8 for class II were not considered due to their minimal effect on binding [17, 19, 43]. For each pocket, contact residues (pocket sequences) were defined as either (1) SLA-specific contacts (Ssc) derived from SLA crystal structures or (2) HLA-based contacts (Hc) derived from HLA crystal structures (Fig. 2a). The Ssc approach was applied only to SLA class I alleles using crystallographic data available for SLA-1*0401 (PDB:3QQ3 and 3QQ4) [2]. For Hc, representative crystal structures from HLA class I and II supertype alleles [41, 42] with bound 9-

Fig. 2 PigMatrix algorithm. **a** Residues in contact with the peptide are determined with respect to a crystal structure of either an SLA allele (Ssc) or an HLA supertype allele (Hc). Class I supertype alleles are represented by three HLA molecules. **b** Contact residues defined by either Ssc or Hc are extracted from HLA sequences and compiled into a library of HLA pockets (HLA pocket library). Pocket A positions and the extracted pocket sequences for Ssc and Hc are marked with *. **c** For a target SLA allele, contact residues (defined by Ssc or Hc) are identified by aligning the SLA sequence to a reference HLA sequence. **d** SLA pocket sequences are compared to those in the HLA pocket library to identify the best human match. **e** Binding coefficients of the best human match for each SLA pocket sequence are compiled to build a PigMatrix. Coefficients are represented in red to blue scale (high to low binding likelihood)

Gutiérrez *et al. BMC Bioinformatics* (2015) 16:290

Page 5 of 13

mer (for class I) or longer peptides (for class II) and the highest resolution were analyzed to define the contact residues (Additional file 2). Four class II supertype alleles (HLA-DRB1*0101, 0301, 0401, and 1501) had crystal structures available.

We considered binding pocket contact residues to be amino acids with atoms within 5.0 Å of those in the bound peptide. Residues were selected using PyMOL (Schrodinger, LLC). Only amino acids with the side chain oriented towards the peptide were included. Thus, each pocket included the union of contact residues in all the MHC crystal structures of a given class. For class II alleles, since the alpha subunit of HLA-DR (HLA-DRA) is practically invariable [44], only residues in the beta subunit were included. The amino acids in the positions defined as contact residues according to Ssc and Hc were extracted from HLA sequences and compiled into a pocket library (Fig. 2b), where each pocket is a non-contiguous sequence of residues ordered by their positions.

Composite matrix construction

Each SLA protein sequence was aligned to a reference HLA sequence (HLA-A*0101 for class I and HLA-DRB1*0101 for class II) to extract its contact residues (Fig. 2c) based on Ssc and Hc approaches. SLA pockets were compared to the HLA pocket library to identify the best human match for each SLA pocket. SLA-HLA pocket similarity was determined using PAM and BLOSUM substitution matrices for closely (PAM30 and BLOSUM90) and distantly related (PAM120 and BLOSUM62) protein sequences [45, 46]. For a pocket comparison between SLA sequence $\mathbf{x}$ and corresponding HLA sequence $\mathbf{y}$, both of length N, the similarity score was calculated as the sum of the similarity scores of each amino acid i using a specific substitution matrix M. The score was then divided by the similarity score of the SLA pocket compared to itself.

$$\text{sim}_M(\mathbf{x},\mathbf{y}) = \frac{\sum_{i=1}^{N} M[x_i,,y_i]}{\sum_{i=1}^{N} M[x_i,,y_i]}$$

The HLA pocket with the highest similarity score was considered to be the best human match (Fig. 2d). The pocket profiles of the best human matches for each pocket were then combined to form composite matrices (Fig. 2e).

Matrix validation and performance evaluation

Composite matrices were used to score a set of random 9-mer peptides. The raw binding score $\text{bind}_{\text{raw}}(\mathbf{p})$ for each peptide $\mathbf{p}$ was calculated as the sum, over a set I of relevant peptide positions, of the coefficient $K[i, p_i]$ of

the amino acid p_i at position i in $\mathbf{p}$. Positions 1, 2, 3, 6, 7, and 9 were used for class I and 1, 4, 6, 7, and 9 for class II, as those positions most interact with each SLA pocket.

$$\text{bind}_{\text{raw}}(\mathbf{p}) = \sum_{i \in l} K[i, p_i]$$

The average μ and standard deviation σ of the scores were used to normalize scores into a Z-score scale (binding likelihood score).

$$Z = \frac{\text{bind}_{\text{raw}}(\mathbf{p}) - \mu}{\sigma}$$

Next, the ability of the composite matrices to separate binders from non-binders and binders from a set of random peptides was evaluated by comparing the mean of the Z-scores of the datasets [28]. Differences were evaluated for significance by Wilcoxon-Mann–Whitney test. For class II binders longer than 10 amino acids, 9-mers overlapping by eight amino acids were scored because in general, the lengths of MHC binding cores are 9 amino acids [47]. The 9-mer frame with the highest Z-score was selected to be the most likely MHC binder and its score was used for calculation of the mean Z-score of binders and non-binders. In addition, for each allele, the HLA matrix with the lowest overall pocket identity with the SLA was used to score both set of peptides, binders and random peptides, as a negative control matrix.
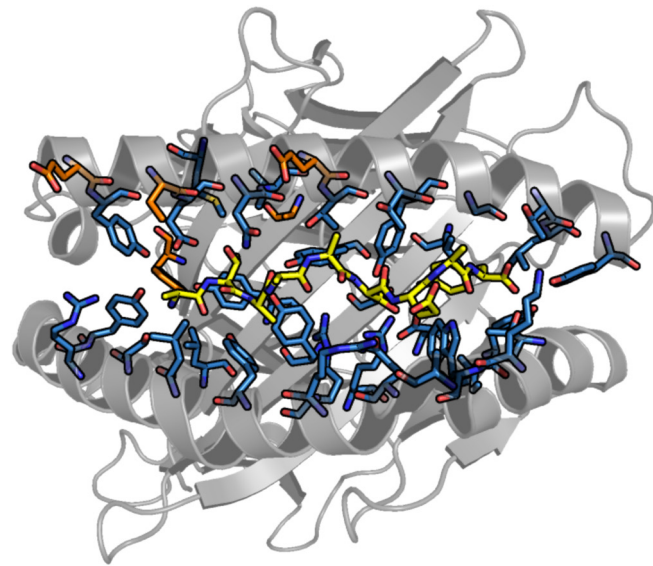
Peptides in the top 5 % of the normal curve, where the Z-score is greater than or equal to 1.64, were considered to be potential binders. This threshold has been shown to identify peptides that are highly likely to bind HLA molecules [22]. So as to evaluate the predictive performance of the matrices, we calculated the area under the receiver operating characteristic (ROC) curve (AUC) using the sensitivity and 1 - *specificity* values for the same dataset of binders and non-binders.

Finally, PigMatrix SLA class I predictions were compared to those of PickPocket 1.1 and NetMHCpan 2.4 and 2.8. A threshold of 500 nM in binding affinity (or 0.426 prediction score based on $1 – log_{50K}(affinity)$), was set to classify binders and non-binders as previously described [6, 18].

Results
Pocket residues

The contact residues that form the binding pockets in SLA were defined from (1) SLA-1*0401 crystal structures (SLA-specific contacts, Ssc), and (2) HLA class I and II crystal structures (HLA contacts, Hc). Figure 3 shows contact residue similarities and differences for SLA-1*0401 in determinations using Hc and Ssc. Thirty-nine positions were identified with Hc, of which 34 were in common with Ssc (shown in light blue in Fig. 3) and

Gutiérrez *et al. BMC Bioinformatics* (2015) 16:290

Page 6 of 13

Fig. 3 (See legend on next page.)

Gutiérrez *et al. BMC Bioinformatics* (2015) 16:290

Page 7 of 13

(See figure on previous page.)

Fig. 3 Comparison of contact residues in the binding pockets based on SLA-specific contacts (Ssc) and HLA contacts (Hc). **Left:** Schematic representation of the crystal structure of SLA-1*0401 (PDB:3QQ4; residues 1 to 181 rendered with PyMOL (Schrodinger, LLC)) showing the residues involved in the binding pockets. SLA contact residues and the ligand (ATAAATEAY, yellow) are represented as sticks. Residues common for both Hc and Ssc approaches are show in light blue; residues unique to Ssc in green and unique to Hc in orange. **Right:** Positions in the SLA binding pockets are shown. The first column (SLA position) is the residue and position number in the SLA-1*0401 protein sequence (Genbank:2352988). Residue positions shown in bold and underlined are identical (*i.e.* amino acid involved in the same pocket(s)) for both approaches. Positions in light blue are common for both approaches; positions in orange are unique to Hc. The next columns show, shaded in gray, the positions involved in pockets A through F that interact with relative ligand positions (peptide position). The last column (Count) is the number of pockets in which an amino acid participates. The last row (Total) is the total number of residues in each pocket

five were unique to Hc (in orange); there were no positions unique to Ssc. Several amino acids were involved in more than one pocket; however, this was more frequent for Hc than Scc due to the nature of the approach; only 23 of the 34 common positions belonged to exactly the same pockets by both definitions (positions shown in bold and underlined in Fig. 3). Hc included for each pocket, the union of amino acids over all the HLA crystal structures analyzed. Based on Hc, positions 97, 99 and 114 were part of four pockets; these residues are located in the central part of the MHC binding groove and depending on the characteristics of their R chain and the bound peptide, they can interact with more than one residue of the ligand. We also observed that the total number of contact residues per pocket was lower in Ssc. The main differences in SLA class I were observed for pockets C and D where SLA structures had fewer contact residues involved in the binding. Pocket F, on the other hand, was identical for both.

For class II, only Hc was applied because no SLA-DR crystal structures were available. Twenty-two positions were considered in contact with ligands; four positions were common to three pockets, four to two pockets and 14 were involved in only one pocket. Five positions were included for pocket A, seven in B, seven in C, eight in D, and seven in F (Additional file 3). In pockets B, C, D, and E, we identified in total 8 allele-specific pocket residues; Y30 in pocket B for HLA-DRB1*1501, D28 and R74 in pocket C for DRB1*0301, Q70 and R74 for DRB1*0301 and R13 for DRB1*1501 in pocket D, and V38 for DRB1*0301 and Y37 for DRB1*0401 in pocket E.

Matrix construction, validation and evaluation

We built composite matrices for each SLA allele and evaluated whether they were able to distinguish SLA allele-specific binders from non-binders and random peptides (Fig. 4). In some cases, the best human matches were the same regardless of the approach used to define pocket residues and the pocket similarity scoring method applied; therefore, prediction results were identical (e.g. Fig. 4 left, SLA-2*0401). For SLA-1*0401 and 2*0401, two and four scoring methods respectively, generated Ssc matrices capable of separating binders from non-binders (highlighted in gray in Fig. 4 left). For these matrices, mean Z-scores of binders were above the threshold to be considered a potential binder (1.64) and non-binder Z-scores were below. Furthermore, the difference between the sets of peptides was statistically significant ($p < 0.001$) using a Wilcoxon-Mann–Whitney test. None of the class I Hc matrices was able to distinguish with statistical significance binders from non-binders. Likewise, for all SLA-3*0401 matrices, mean Z-scores of the binders were either not above the 1.64 threshold or the non-binders had higher mean Z-scores. For class II allele SLA-DRB1*0201, binders scored using Hc matrices were above the threshold and were statistically distinct from mean Z-scores of the non-binders ($p < 0.01$). Negative control-matrices for all SLA alleles (using HLA alleles with the lowest overall pocket identity) did not separate binders from random peptides (means range from −0.57 to 1.13), showing that the selection of the best human match based on similarity is critical. In sum, these results show that some composite class I Ssc matrices and class II Hc were able to separate binders from random peptides and non-binders.

To evaluate the predictive performance of the matrices, we built ROC curves and then calculated the AUC. For the AUC, a value of 0.5 corresponds to a random prediction and a value of 1 to a perfect prediction. Figure 5 shows a comparison of the AUCs of Ssc and Hc matrices, PickPocket and NetMHCpan. For class I, matrices based on Ssc had higher AUCs than Hc-based matrices. Class I and II matrices constructed using PAM30 to score pocket similarity had higher AUCs compared to matrices constructed using PAM120, BLOSUM62 and PAM120, with one exception (Hc SLA-1*0401 built using BLOSUM62). Compared to PickPocket and NetMHCpan 2.4 and 2.8, PigMatrix's AUC was equivalent or better for SLA-1*0401 and 2*0401; however, due to the nature of the tests, we could not assess statistical significance. It is worth noting that, in contrast with NetMHCpan 2.8, SLA peptide data did not contribute to training PigMatrix..

The SLA-3*0401 PAM30-Ssc matrix had the lowest AUC (0.60). This was not unexpected as matrices for this allele were unable to separate binders from non-

Gutiérrez *et al. BMC Bioinformatics* (2015) 16:290

Page 8 of 13

SLA allele	Similarity score	Mean binding likelihood expressed as Z-score (sd)			
		Ssc		Hc	
		Binders	Non-binders	Binders	Non-binders
1*0401	PAM30	**2.12 (0.46)**	**1.49 (0.57)**	1.65 (0.55)	1.79 (0.48)
	BLOSUM90	2.98 (1.00)	3.06 (1.00)	2.99 (0.96)	3.15 (0.97)
	PAM120	**2.12 (0.46)**	**1.49 (0.57)**	1.61 (0.54)	1.69 (0.51)
	BLOSUM62	2.98 (1.00)	3.06 (1.00)	3.02 (0.97)	3.17 (0.96)
	Negative	-0.58 (0.76)	-0.45 (0.78)	0.51 (0.67)	0.53 (0.52)
2*0401	PAM30	**2.30 (0.76)**	**1.45 (0.84)**	2.07 (0.83)	1.83 (1.10)
	BLOSUM90	**2.30 (0.76)**	**1.45 (0.84)**	2.07 (0.83)	1.83 (1.10)
	PAM120	**2.30 (0.76)**	**1.45 (0.84)**	2.07 (0.83)	1.83 (1.10)
	BLOSUM62	**2.30 (0.76)**	**1.45 (0.84)**	2.20 (0.98)	1.96 (1.09)
	Negative	**0.56 (0.59)**	**1.00 (0.75)**	0.56 (0.59)	1.00 (0.75)
3*0401	PAM30	1.42 (1.05)	1.16 (0.77)	1.34 (1.17)	1.12 (0.78)
	BLOSUM90	**1.62 (0.81)**	**2.20 (0.82)**	1.71 (0.90)	1.96 (0.82)
	PAM120	**1.36 (0.72)**	**1.95 (0.78)**	1.27 (1.06)	1.68 (0.95)
	BLOSUM62	**1.74 (0.79)**	**2.22 (0.75)**	1.98 (0.90)	2.32 (0.87)
	Negative	0.52 (0.70)	0.47 (0.55)	**0.54 (0.82)**	**1.13 (0.96)**
DRB1*0201	PAM30			**1.85 (0.65)**	**1.29 (0.52)**
	BLOSUM90			**1.67 (0.57)**	**1.36 (0.55)**
	PAM120			**1.67 (0.57)**	**1.36 (0.55)**
	BLOSUM62			**1.67 (0.57)**	**1.36 (0.55)**
	Negative			**0.30 (0.42)**	**-0.03 (0.43)**

Fig. 4 (See legend on next page.)

Gutiérrez *et al. BMC Bioinformatics* (2015) 16:290

Page 9 of 13

binders as described above (Fig. 4). For these reasons, we examined the binding preferences of the best human matches for the PAM30-Ssc matrix and compared them to the amino acid frequencies in the sets of binders and non-binders. The most evident differences were observed in pocket B. The best human match for pocket B was HLA-A*0301 ($\text{sim}_{\text{PAM30}}$ 0.49). Of all binders reported for SLA-3*0401, the most common residue in position two (Pocket B) was arginine, found in 37 % of binders, followed by alanine, found in 19 % of binders. These frequencies were more similar to the binding preferences in HLA-B*2705 ($\text{sim}_{\text{PAM30}}$ 0.17) pocket profile, in contrast to A*0301, which had negative coefficients for arginine and alanine. Based on this observation, we modified pocket B of PAM30-Ssc matrix from A*0301 to B*2705 (Ssc_{ModB}). This matrix had a higher AUC (0.81) than the original matrix (Fig. 5). This result showed that predictions could potentially be improved by selecting best human matches based on similarity in terms of binding preferences if

binding information is available. However, it is important to note that this improvement in the AUC was specific to the set of peptides available to date, and further prospective studies are needed to validate the preference of this particular pocket.

Overall, these results demonstrate that matrices built using contact residues from SLA structures and using the PAM30 substitution matrix to identify the best human match for each pocket, had the best predictive power of the approaches that were tested. Although the matrices showed predictive power, the limited number of known binders makes the AUC values less robust. For this reason, an analysis of a larger dataset of SLA-specific binders and non-binder peptides will be required to revalidate the predictive power of the matrices.

Discussion and conclusions

Immunoinformatics tools have accelerated the identification of epitopes and design of human vaccines. However,
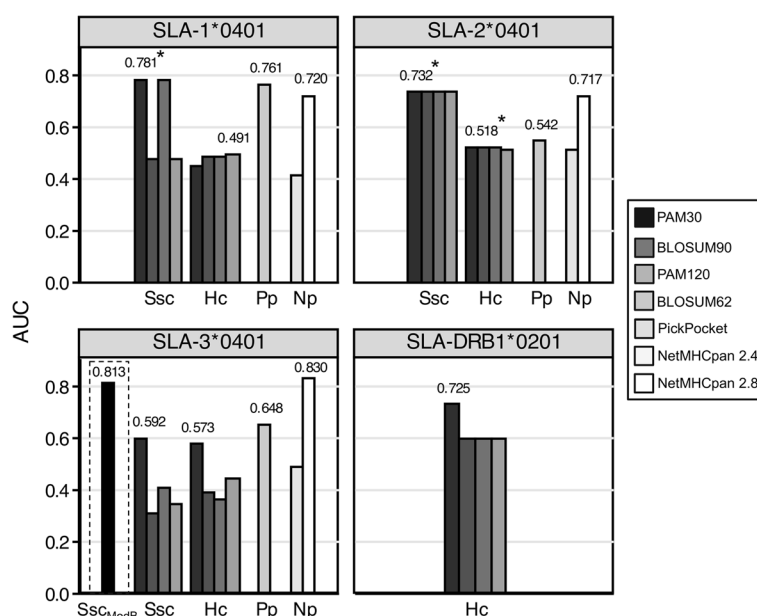

Fig. 5 Matrix performance comparison. AUCs of the matrices built for SLA class I and II alleles are shown. The highest AUC for each method is shown above the bars. For SLA-3*0401, AUC of Ssc_{ModB}, a PAM30-Ssc matrix with a pocket B profile different than the best human match, is shown in a dashed rectangle to illustrate the impact of pocket B. If one or more matrices for the same approach (Ssc or Hc) have equal AUC, it is indicated with *

Gutiérrez *et al. BMC Bioinformatics* (2015) 16:290

Page 10 of 13

comparable tools have not been applied extensively to pigs. During the last two decades, swine T cell epitope discovery has been based on experimental studies of numerous overlapping peptides [37–39, 48–54]. While these studies are essential for validating T cell epitope prediction tools, they can be expensive and time consuming. To reduce experimental effort and expedite the process, algorithms developed for human T cell epitope prediction have also been used to identify porcine epitopes [55–57]. However, the substitution of HLA predictions for SLA predictions may not reflect the fine specificity of SLA binding, which limits the efficacy of this oversimplified approach. To overcome this, we have developed PigMatrix, a simple yet effective method that leverages available data (SLA-binding peptides, SLA structures and HLA binding data) and pocket profiles already constructed for HLA-based epitope prediction in EpiMatrix to predict potential T cell epitopes for SLA class I and II alleles. Using the pocket profile method and the concept that pockets that have similar amino acids will share similar binding preferences, we built and validated matrices that were able to separate SLA-restricted peptides from random peptides and non-binders.

Human pan-specific tools based on the pocket profile method have been described for prediction of class I (PickPocket) and class II T cell epitopes (TEPITOPEpan) [9, 18]. These methods defined the amino acids in the pockets from HLA crystal structures. Similarly, we defined SLA-specific contacts (Ssc) from two crystal structures available for SLA-1*0401. Additionally, we extrapolated the pocket residues from crystal structures of HLA class I and II crystal structures (Hc). Both approaches assume for a given pocket that all contact residues are conserved across all class-specific MHC molecules. However, because there are differences in the pocket residues between MHC alleles and even between the same allele structures depending on the ligand [2], this simplification is a limitation of the peptide:MHC modeling approach. Even so, it is a reasonable approximation when structural information is limited. For class I, differences in the pocket residues using the Hc and the Ssc approaches were noticeable and impacted the subsequent selection of the best human match to build the prediction matrices. Matrices based on pockets defined from SLA structure-specific contacts performed better than HLA-derived pockets. While a definition of the contact residues based on several HLA structures account for the intra- and inter-allelic variability of binding pockets, it also dilutes the importance of key residues in the peptide:MHC interaction. We speculate that more allele-specific pockets could potentially improve the selection of the best human match and therefore the predictions. Selection might be also improved

by weighting the similarity score by conservation of key contact residues.

PickPocket and TEPITOPEpan use a method based on BLOSUM62 to calculate a weighted score of specificity to define the most similar HLA-derived pocket. For PigMatrix, in addition to BLOSUM62, we used PAM120, which is considered equivalent to BLOSUM62 for comparison of distantly related proteins [46], to calculate pocket similarity. We also included PAM30 and BLOSUM90, which are both designed to score similarity between closely related protein sequences. The SLA matrices with highest AUC were based on PAM30 using both Hc and Ssc, with only one exception. If we consider the pocket contact residues as short pseudo-sequences, the better performance of PAM30-based matrices might be explained because low-numbered PAM matrices are more efficient for searches involving short sequences. BLOSUM62 on the other hand, performs better to identify distant homologs using longer sequences. BLOSUM90, like PAM30, is used for closely related sequences; however, it is not recommended for short peptides [46] and unsurprisingly did not perform as well as PAM30 in these studies.

Predictive methods for porcine T-cell epitopes are limited, and none existed previously for SLA class II. PigMatrix is the first tool to make binding predictions for an SLA-DR allele. Class II predictions were limited to the Hc method because no SLA-DR molecule has been crystallized. Since SLA-DR-specific binding data are scarce, predictions require further prospective validation. While it is not possible at this time to benchmark the SLA-DR matrix against other predictors, a comparison can be made for the SLA class I matrices developed here. NetMHCpan has been used for SLA binding predictions. PickPocket, which is also based on the pocket profile method, has been described primarily for HLA class I [18], but predictions are also available for SLA alleles. In this study, for an existing set of published peptides, PigMatrix performed equally or better than two versions of NetMHCpan and PickPocket for SLA-1*0401 and 2*0401. While PigMatrix and PickPocket derive SLA binding preferences from HLA binding data, NetMHCpan artificial neural networks are trained using information derived from available binding data as well as peptide sequences and MHC sequence information [6]. It was previously demonstrated that in a scenario where the quantitative binding data were limited for human and non-human MHC alleles, PickPocket performed better than NetMHCpan [18]. This is also evident when NetMHCpan 2.4 results are compared to NetMHCpan 2.8 predictions. NetMHCpan 2.4 was trained with a limited set of SLA binders and its predictions were equivalent to random selection (average AUC 0.47). Version 2.8, on the other hand, was trained with more data and

its performance improved for the alleles we evaluated (average AUC 0.76). Conversely, PigMatrix was not trained with SLA-specific binding data and performed similarly or better than NetMHCpan 2.8 predictions for two of three class I alleles we tested. Moreover, because the number of published peptides is limited, we were not able to compile a test dataset of peptides known to be different from the training set used by NetMHCpan 2.8. Hence, it is possible that NetMHCpan 2.8 performance was overestimated.

For SLA-3*0401, PickPocket and NetMHCpan 2.8 outperformed PigMatrix. Upon closer analysis, these results provided an example of how PigMatrix could be improved. We were able to build a better performing model by modifying pocket B in the matrix constructed using PAM30-Ssc. This might be explained by the role of the amino acid in position two of the peptide as a binding anchor and its specific interaction with this pocket. It is also worth noting that the HLA pocket library we used was limited to 12 class I and eight class II supertypes alleles available in EpiMatrix. It is possible that pocket sequences from other HLA alleles and their profiles are more similar to SLA pockets. Therefore, if the number of HLA alleles in the library is increased, we might find better human matches for SLA pockets, which could potentially improve matrix performance.

So as to illustrate the PigMatrix approach, we built initially matrices for only three SLA class I alleles and one class II SLA allele for which quantitative binding data were available. These alleles are commonly expressed in different porcine breeds and cell lines for *in vitro* culture [32–36]. However, like HLA, SLA diversity is considerable. These results demonstrate the potential of the approach to be extended to SLA alleles with limited or nonexistent epitope binding data. Thus, future versions of PigMatrix will include a more comprehensive and representative set of matrices for SLA alleles expressed in outbred porcine populations. Moreover, prospective *in vitro* and *in vivo* evaluation of PigMatrix predictions will help to refine the matrices.

We developed the PigMatrix tool with the intent to integrate it into the iVAX toolkit, which is a comprehensive set of tools for computational vaccine design that includes EpiMatrix, Conservatrix, ClustiMer, EpiAssembler, JanusMatrix, and VaccineCAD [58]. When the PigMatrices are used to substitute for HLA matrices (EpiMatrix) in iVAX, all of the existing suite of iVAX vaccine design tools can be used with the SLA epitope predictions, which makes it possible to envision accelerated development of novel T cell epitope-based vaccines or whole subunit vaccines optimized for epitope content that protect against infectious disease in swine.

Additional files

> **Additional file 1: Peptide database.** Exp: Experimental results; 0: Non-binders, 1: Binder.(DOCX 156 kb)
>
> **Additional file 2: HLA crystal structures.**
>
> **Additional file 3: Contact residues in the SLA class II binding pockets based on HLA contacts (Hc).**

Author details

[1]Institute for Immunology and Informatics, CMB/CELS, University of Rhode Island, Providence, RI 02903, USA. [2]EpiVax, Inc., Providence, RI 02860, USA. [3]Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA.

References

1. Neefjes J, Jongsma MLM, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol. 2011;11:823–36.

Gutiérrez *et al. BMC Bioinformatics* (2015) 16:290

Page 12 of 13

2. Zhang N, Qi J, Feng S, Gao F, Liu J, Pan X, et al. Crystal structure of swine major histocompatibility complex class I SLA-1 0401 and identification of 2009 pandemic swine-origin influenza A H1N1 virus cytotoxic T lymphocyte epitope peptides. J Virol. 2011;85:11709–24.

3. Wee LJK, Lim SJ, Ng LFP, Tong JC. Immunoinformatics: how in silico methods are re-shaping the investigation of peptide immune specificity. Front Biosci Elite Ed. 2012;4:311–9.

4. Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, et al. Immune epitope database analysis resource (IEDB-AR). Nucleic Acids Res. 2008;36(Web Server issue):W513–8.

5. Jojic N, Reyes-Gomez M, Heckerman D, Kadie C, Schueler-Furman O. Learning MHC I–peptide binding. Bioinforma Oxf Engl. 2006;22:e227–35.

6. Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS ONE. 2007;2, e796.

7. Jacob L, Vert J-P. Efficient peptide-MHC-I binding prediction for alleles with few known binders. Bioinforma Oxf Engl. 2008;24:358–66.

8. Bordner AJ, Mittelmann HD. MultiRTA: a simple yet reliable method for predicting peptide binding affinities for multiple class II MHC allotypes. BMC Bioinformatics. 2010;11:482.

9. Zhang L, Chen Y, Wong H-S, Zhou S, Mamitsuka H, Zhu S. TEPITOPEpan: Extending TEPITOPE for Peptide Binding Prediction Covering over 700 HLA-DR Molecules. PLoS ONE. 2012;7, e30483.

10. Guo L, Luo C, Zhu S. MHC2SKpan: a novel kernel based approach for pan-specific MHC class II peptide binding prediction. BMC Genomics. 2013;14 Suppl 5:S11.

11. Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. Immunogenetics. 2013;65:711–24.

12. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics. 2009;61:1–13.

13. Pedersen LE, Harndahl M, Rasmussen M, Lamberth K, Golde WT, Lund O, et al. Porcine major histocompatibility complex (MHC) class I molecules and analysis of their peptide-binding specificities. Immunogenetics. 2011;63:821–34.

14. Pedersen LE, Jungersen G, Buus S, Golde WT: Analysis of Swine Leukocyte Antigen peptide binding profiles and the identification of T cell epitopes by tetramer staining. Ph.D. thesis. Technical University of Denmark; 2012.

15. Pedersen LE, Harndahl M, Nielsen M, Patch JR, Jungersen G, Buus S, et al. Identification of peptides from foot-and-mouth disease virus structural proteins bound by class I swine leukocyte antigen (SLA) alleles, SLA-1*0401 and SLA-2*0401. Anim Genet. 2013;44:251–8.

16. Pedersen LE, Breum SO, Riber U, Larsen LE, Jungersen G. Identification of swine influenza virus epitopes and analysis of multiple specificities expressed by cytotoxic T cell subsets. Virol J. 2014;11:163.

17. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol. 1999;17:555–61.

18. Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. Bioinformatics. 2009;25:1293–9.

19. Chelvanayagam G. A roadmap for HLA-A, HLA-B, and HLA-C peptide binding specificities. Immunogenetics. 1996;45:15–26.

20. Chelvanayagam G. A Roadmap for HLA-DR Peptide Binding Specificities. Hum Immunol. 1997;58:61–9.

21. Rammensee H-G, Bachmann J, Stevanovic S. MHC Ligands and Peptide Motifs. Georgetown, TX: Landes Bioscience; 1997.

22. De Groot AS, Jesdale BM, Szu E, Schafer JR, Chicz RM, Deocampo G. An interactive Web site providing major histocompatibility ligand predictions: application to HIV research. AIDS Res Hum Retroviruses. 1997;13:529–31.

23. McMurry JA, Kimball S, Lee JH, Rivera D, Martin W, Weiner DB, et al. Epitope-driven TB vaccine development: a streamlined approach using immuno-informatics, ELISpot assays, and HLA transgenic mice. Curr Mol Med. 2007;7:351–68.

24. Gregory SH, Mott S, Phung J, Lee J, Moise L, McMurry JA, et al. Epitope-based vaccination against pneumonic tularemia. Vaccine. 2009;27:5299–306.

25. Moise L, Buller RM, Schriewer J, Lee J, Frey SE, Weiner DB, et al. VennVax, a DNA-prime, peptide-boost multi-T-cell epitope poxvirus vaccine, induces protective immunity against vaccinia infection by T cell response alone. Vaccine. 2011;29:501–11.

26. Moss SF, Moise L, Lee DS, Kim W, Zhang S, Lee J, et al. HelicoVax: epitope-based therapeutic Helicobacter pylori vaccination in a mouse model. Vaccine. 2011;29:2085–91.

27. De Groot AS, Martin W. Reducing risk, improving outcomes: bioengineering less immunogenic protein therapeutics. Clin Immunol. 2009;131:189–201.

28. De Groot AS, Jesdale B, Martin W, Saint Aubin C, Sbai H, Bosma A, et al. Mapping cross-clade HIV-1 vaccine epitopes using a bioinformatics approach. Vaccine. 2003;21:4486–504.

29. Moise L, McMurry JA, Buus S, Frey S, Martin WD, De Groot AS. In silico-accelerated identification of conserved and immunogenic variola/vaccinia T-cell epitopes. Vaccine. 2009;27:6471–9.

30. Mishra S, Losikoff PT, Self AA, Terry F, Ardito MT, Tassone R, et al. Peptide-pulsed dendritic cells induce the hepatitis C viral epitope-specific responses of naïve human T cells. Vaccine. 2014;32:3285–92.

31. De Groot AS, Nene V, Hegde NR, Srikumaran S, Rayner J, Martin W. T cell epitope identification for bovine vaccines: an epitope mapping method for BoLA A-11. Int J Parasitol. 2003;33:641–53.

32. Smith DM, Lunney JK, Martens GW, Ando A, Lee J-H, Ho C-S, et al. Nomenclature for factors of the SLA class-I system, 2004. Tissue Antigens. 2005;65:136–49.

33. Smith DM, Lunney JK, Ho C-S, Martens GW, Ando A, Lee J-H, et al. Nomenclature for factors of the swine leukocyte antigen class II system, 2005. Tissue Antigens. 2005;66:623–39.

34. Ho C-S, Franzo-Romain MH, Lee YJ, Lee JH, Smith DM. Sequence-based characterization of swine leucocyte antigen alleles in commercially available porcine cell lines. Int J Immunogenet. 2009;36:231–4.

35. Ho C-S, Lunney JK, Franzo-Romain MH, Martens GW, Lee Y-J, Lee J-H, et al. Molecular characterization of swine leucocyte antigen class I genes in outbred pig populations. Anim Genet. 2009;40:468–78.

36. Ho C-S, Lunney JK, Lee J-H, Franzo-Romain MH, Martens GW, Rowland RRR, et al. Molecular characterization of swine leucocyte antigen class II genes in outbred pig populations. Anim Genet. 2010;41:428–32.

37. Ober BT, Summerfield A, Mattlinger C, Wiesmüller KH, Jung G, Pfaff E, et al. Vaccine-induced, pseudorabies virus-specific, extrathymic CD4 + CD8+ memory T-helper cells in swine. J Virol. 1998;72:4866–73.

38. Armengol E, Wiesmüller K-H, Wienhold D, Büttner M, Pfaff E, Jung G, et al. Identification of T-cell epitopes in the structural and non-structural proteins of classical swine fever virus. J Gen Virol. 2002;83:551–60.

39. Gerner W, Denyer MS, Takamatsu H-H, Wileman TE, Wiesmüller K-H, Pfaff E, et al. Identification of novel foot-and-mouth disease virus specific T-cell epitopes in c/c and d/d haplotype miniature swine. Virus Res. 2006;121:223–8.

40. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31:3784–8.

41. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. BMC Immunol. 2008;9:1.

42. Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, et al. Several common HLA-DR types share largely overlapping peptide binding repertoires. J Immunol. 1998;160:3363–73.

43. Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, et al. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature. 1994;368:215–21.

44. Stern LJ, Calvo-Calle JM. HLA-DR: molecular insights and vaccine design. Curr Pharm Des. 2009;15:3249–61.

45. Dayhoff MO, Schwartz RM. A model of evolutionary change in proteins. In: Atlas of protein sequence and structure. 1978.

46. Altschul SF. Amino acid substitution matrices from an information theoretic perspective. J Mol Biol. 1991;219:555–65.

47. Madden DR. The three-dimensional structure of peptide-MHC complexes. Annu Rev Immunol. 1995;13:587–622.

Gutiérrez *et al. BMC Bioinformatics* (2015) 16:290

Page 13 of 13

48. Blanco E, McCullough K, Summerfield A, Fiorini J, Andreu D, Chiva C, et al. Interspecies major histocompatibility complex-restricted Th cell epitope on foot-and-mouth disease virus capsid protein VP4. J Virol. 2000;74:4902–7.

49. Blanco E, Garcia-Briones M, Sanz-Parra A, Gomes P, De Oliveira E, Valero ML, et al. Identification of T-cell epitopes in nonstructural proteins of foot-and-mouth disease virus. J Virol. 2001;75:3164–74.

50. García-Briones MM, Blanco E, Chiva C, Andreu D, Ley V, Sobrino F. Immunogenicity and T cell recognition in swine of foot-and-mouth disease virus polymerase 3D. Virology. 2004;322:264–75.

51. Stevenson LS, Gilpin DF, Douglas A, McNeilly F, McNair I, Adair BM, et al. T lymphocyte epitope mapping of porcine circovirus type 2. Viral Immunol. 2007;20:389–98.

52. Vashisht K, Goldberg TL, Husmann RJ, Schnitzlein W, Zuckermann FA. Identification of immunodominant T-cell epitopes present in glycoprotein 5 of the North American genotype of porcine reproductive and respiratory syndrome virus. Vaccine. 2008;26:4747–53.

53. Wang Y-X, Zhou Y-J, Li G-X, Zhang S-R, Jiang Y-F, Xu A-T, et al. Identification of immunodominant T-cell epitopes in membrane protein of highly pathogenic porcine reproductive and respiratory syndrome virus. Virus Res. 2011;158:108–15.

54. Parida R, Choi I-S, Peterson DA, Pattnaik AK, Laegreid W, Zuckermann FA, et al. Location of T-cell epitopes in nonstructural proteins 9 and 10 of type-II porcine reproductive and respiratory syndrome virus. Virus Res. 2012;169:13–21.

55. Díaz I, Pujols J, Ganges L, Gimeno M, Darwich L, Domingo M, et al. In silico prediction and *ex vivo* evaluation of potential T-cell epitopes in glycoproteins 4 and 5 and nucleocapsid protein of genotype-I (European) of porcine reproductive and respiratory syndrome virus. Vaccine. 2009;27:5603–11.

56. Zimic M, Gutiérrez AH, Gilman RH, López C, Quiliano M, Evangelista W, et al. Immunoinformatics prediction of linear epitopes from Taenia solium TSOL18. Bioinformation. 2011;6:271–4.

57. Burgara-Estrella A, Díaz I, Rodríguez-Gómez IM, Essler SE, Hernández J, Mateu E. Predicted peptides from non-structural proteins of porcine reproductive and respiratory syndrome virus are able to induce IFN-γ and IL-10. Viruses. 2013;5:663–77.

58. Terry FE, Moise L, Martin RF, Torres M, Pilotte N, Williams SA, et al. Time for T? Immunoinformatics addresses vaccine design for neglected tropical and emerging infectious diseases. Expert Rev Vaccines. 2015;14:21–35.