Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2007) 667–675
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Structure-based prediction of MHC–peptide association: Algorithm comparison and application to cancer vaccine design

Alexandra J. Schiewe, Ian S. Haworth *

Department of Pharmacology & Pharmaceutical Sciences, University of Southern California, 1985 Zonal Avenue, Los Angeles, CA 90089-9121, USA

Accepted 30 March 2007
Available online 4 April 2007

Abstract

Peptide vaccination for cancer immunotherapy requires identification of peptide epitopes derived from antigenic proteins associated with the tumor. Such peptides can bind to MHC proteins (MHC molecules) on the tumor-cell surface, with the potential to initiate a host immune response against the tumor. Computer prediction of peptide epitopes can be based on known motifs for peptide sequences that bind to a certain MHC molecule, on algorithms using experimental data as a training set, or on structure-based approaches. We have developed an algorithm, which we refer to as PePSSI, for flexible structural prediction of peptide binding to MHC molecules. Here, we have applied this algorithm to identify peptide epitopes (of nine amino acids, the common length) from the sequence of the cancer-testis antigen KU-CT-1, based on the potential of these peptides to bind to the human MHC molecule HLA-A2. We compared the PePSSI predictions with those of other algorithms and found that several peptides predicted to be strong HLA-A2 binders by PePSSI were similarly predicted by another structure-based algorithm, PREDEP. The results show how structure-based prediction can identify potential peptide epitopes without known binding motifs and suggest that side chain orientation in binding peptides may be obtained using PePSSI.

Keywords: MHC; HLA-A2; KU-CT-1; Peptide; Epitope; Structure prediction

1. Introduction

The goal of cancer immunotherapy is activation of a host immune response against tumor cells [1,2]. Current approaches include vaccination with dendritic cells harvested from a patient and activated directly with a tumor antigen or via transfection with a gene for the antigen [3]; T-cell adoptive immunotherapy, in which a T-cell population reactive to the tumor is expanded in vitro and then transferred to the patient [4]; and peptide vaccination, in which an antigenic peptide is administered with the goal of eliciting specific T-cell reactivity to tumors [5,6]. The basic aim of each approach is to increase the interaction between cytotoxic T-cells and the complex formed between an antigenic peptide and a major histocompatability complex protein (MHC molecule) on the tumor-cell surface, since engagement of the T-cell receptor (TCR) with the peptide–MHC complex is the first step in elicitation of an immune response [7].

MHC molecules are cell surface glycoproteins that present antigen-derived peptides to T-cells, eliciting an immune response [7]. Class I MHC molecules, which are the focus of the current article, present peptides that are processed via the proteasomal pathway to CD8+ T-cells, whereas class II MHC molecules present peptides processed through the lysosomal pathway to CD4+ T-cells. X-ray crystal structures have provided many insights into the binding modes of peptides with MHC molecules [8]; hence, peptides bound to human class I MHC molecules typically contain nine amino acids, and adopt a somewhat extended conformation in which the N-and C-termini of the peptide are buried in the MHC binding groove and the central part of the peptide bulges from the groove [9]. MHC molecules are polymorphic and binding of peptides to a specific MHC molecule shows sequence dependence; this has led to the concept of ‘‘anchor’’ residues at certain positions in the peptide [9]. For example, the most common human MHC molecule, HLA-A2 (human leukocyte antigen A2), is thought

* Corresponding author. Tel.: +1 323 442 3310; fax: +1 323 442 1390.
E-mail address: ihaworth@usc.edu (I.S. Haworth).

1093-3263/$ – see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2007.03.017

to bind favorably to peptides with large hydrophobic side chains at positions 2 and 9 [9,10]; however, as shown below, use of anchor residues in a motif-based prediction approach may not be sufficient to discriminate fully between peptides that do and do not bind to HLA-A2.

For peptide vaccination, the antigenic peptide is derived from a protein that is overexpressed in the targeted tumor cells [11]; one of the best known examples of this is the Gp2 peptide fragment of the HER2/Neu protein, which is commonly overexpressed in breast cancer [12,13]. Therefore, prediction of whether a peptide will bind to a particular human MHC molecule is of importance for development of peptide-based vaccines [11], and computational approaches have been created to identify such peptides. These approaches have included algorithms using primary anchor motifs [14,15], methods based on experimental binding data [16–18], 3D-QSAR [19], and structure-based prediction [20,21]. Several of the more recent algorithms [22–26] incorporating one or more of these approaches are discussed below.

Water molecules are also of importance in mediating the interaction between peptides and MHC molecules [27–29]. We have recently published an algorithm, PePSSI (peptide–MHC prediction of structure through solvated interfaces) that allows flexible peptide structure prediction in the class I MHC binding groove with incorporation of explicit water molecules at the peptide–MHC interface [30]; interface solvation is achieved through a second algorithm, WATGEN, that is embedded in PePSSI and can also be run as a separate program [31]. We have shown that PePSSI-predicted structures of peptide complexes with HLA-A2 match well with X-ray data [30], and we also found that peptide binding affinity with HLA-A2 is positively correlated with peptide–MHC contacts and negatively correlated with the number of interfacial water molecules predicted by PePSSI [30].

Here, we have used PePSSI to examine potential HLA-A2-binding peptides derived from the KU-CT-1 protein, which is a cancer-testis (CT) antigen expressed in pancreatic, lung, and endometrial cancers [32]. CT antigens are attractive targets for immunotherapy since their expression is restricted to germ cells in normal tissue and they have the potential to produce humoral and cellular immune responses [33]. In particular, these antigens are able to drive T-cell responses, suggesting that they are properly degraded by the proteasome into peptides that are able to bind to class I MHC molecules [33]. Using PePSSI, we calculated the structures and binding affinity of all 864 peptides of 9 amino acids in the KU-CT-1 sequence, and we compared these results with data from six other MHC–peptide prediction algorithms.

2. Methods

2.1. PePSSI binding score

To determine a simple equation for a PePSSI binding score, the binding conformations of 266 9-amino acid peptides with HLA-A2 were predicted using PePSSI. The properties of the predicted structures were correlated with a published tabulation of experimental binding affinities (IC_{50}) of these peptides with HLA-A2 [34]. Derivation of the equation was based on the results for 261 of these peptides, with omission of 5 peptides with $IC_{50} > 50$ μM. Since we have previously found significant relationships between peptide binding affinity for HLA-A2 with the PePSSI-predicted number of peptide–MHC contacts and the number of interfacial water molecules [30], these structural parameters were chosen for use in the binding affinity equation. The PePSSI-predicted HLA-A2/peptide complexes were analyzed for the number of MHC–peptide atom pairs with a separation within 95–105% of the sum of their VDW radii (nContact) and the number of interfacial water molecules mediating peptide–MHC association through hydrogen bonding (nWat). The basis of these parameters has been described elsewhere: the range of 95–105% of the sum of the VDW radii for nContact reflects a favorable energetic contact [30]; and we have rigorously tested the parameters used to define the water positions in development of the WATGEN algorithm [31]. In the current work, nContact and nWat were determined as averages for 25 PePSSI-predicted structures for a given peptide [30], and regression analysis of the dataset was used to determine a binding score equation. The resulting equation is discussed further below.

2.2. PePSSI prediction of HLA-A2-binding peptides from KU-CT-1

The KU-CT-1 protein exists as two splice variants, KU-CT-1S and KU-CT-1L [32]. The KU-CT-1L sequence (accession # NP_775104 in the PubMed Protein Database) was broken into 864 9-amino acid peptides for analysis in PePSSI. The binding affinity of each peptide was assessed in an initial PePSSI run, and those peptides that were found to have a strong or moderate binding affinity for HLA-A2 (see below) were subjected to two further PePSSI runs. Each run was performed with flexibility of 12 HLA-A2 side chains: H70, T73, H74, L81, V95, R97, H114, Y116, K146, Q155, L156, and T163. These side chains were identified as conformationally flexible (based on X-ray structures) in the development of the algorithm [30]. The PePSSI algorithm uses random generation of peptide and MHC conformers through torsional variation [30], and therefore each run gives slightly different structural results; however the broad structural predictions are consistent between runs [30] and we show below that the predicted binding affinities are also consistent from run to run. The PePSSI binding affinity is expressed as the mean ± S.D. over three runs.

2.3. Prediction of HLA-A2-binding peptides from KU-CT-1 using other algorithms

Potential HLA-A2-binding peptides from the KU-CT-1L sequence were also assessed using six other MHC–peptide prediction algorithms: SYFPEITHI [14], PREDEP [22], ProPred-I [23], SMM [24], NetMHC [25] and MHCPred [26]. All these algorithms are available on the Internet, and all were used in their default modes: SYFPEITHI,

http://www.syfpeithi.bmi-heidelberg.com/; ProPred-I, http://www.imtech.res.in/raghava/propred1/; SMM, http://zlab.bu.edu/SMM/; PREDEP, http://bioinfo.md.huji.ac.il/marg/Teppred/mhc-bind/; NetMHC, http://www.cbs.dtu.dk/services/NetMHC/; MHCPred, http://www.jenner.ac.uk/MHCPred/.

3. Results

3.1. PePSSI binding score equation

The compilation of HLA-A2 experimental binding affinities (IC_{50}) of 266 9-amino acid peptides [34] was used as a basis for deriving a binding affinity equation for PePSSI, with omission of 5 peptides with very weak affinity for HLA-A2 ($IC_{50} > 50$ μM). The structures of the remaining 261 peptides complexed with HLA-A2 were predicted using PePSSI and structural data from the predicted complexes were correlated with the experimental data [34]. The number of MHC–peptide atom pairs with a separation of 95–105% of the sum of the VDW radii (nContact) and the number of interfacial water molecules (nWat) were determined by averaging the values for 25 PePSSI-predicted structures for each peptide [30]. Linear regression analysis of the 261-peptide dataset gave the following equations:

$$\mathrm{pIC_{50}} = 7.482 + 0.02727(\mathrm{nContact}) - 0.0221(\mathrm{nWat}) \quad (1)$$

$$\text{Binding score} = 10^{-\mathrm{pIC_{50}}} \times 10^{9} \quad (2)$$

The negative logarithm (base 10) of the binding score (pIC_{50}) is obtained from Eq. (1), and the PePSSI binding score is calculated from Eq. (2), nominally as an IC_{50} value with units of nanomolar. Eq. (1) has a correlation coefficient (R^2) of 0.109

Table 1
Prediction of binding to HLA-A2 for peptides derived from the KU-CT-1L sequence

Method	Sequence position	Sequence	Method score[a]	PePSSI score	PePSSI class[b]
PREDEP	845	EMYVIDLMF	−6.99	13.49	Strong
	794	IFYHRALLF	−6.26	14.35	Strong
	584	LLPLKELCL	−6.08	36.05	Strong
	763	KLPDFSWEL	−6.05	54.54	Moderate
	847	YVIDLMFHP	−6.04	40.25	Strong
	770	ELHISELKF	−5.42	23.29	Strong
SYFPEITHI	209	LLALKTLGV	29	61.33	Moderate
	151	GLEPLIRLL	27	97.81	Weak
	399	GIDPLINLL	27	87.14	Weak
	481	GLEPLVELL	27	127.94	Non-binder
	109	VMNSVIAQL	26	118.41	Non-binder
	192	AIPPILDLL	26	33.61	Strong
ProPred-I	763	KLPDFSWEL	8736.688	54.54	Moderate
	602	LLINSKSYV	650.311	125.52	Non-binder
	840	GLPAPEMYV	382.536	46.03	Strong
	546	KLLNNNLSL	276.643	114.54	Non-binder
	209	LLALKTLGV	118.238	61.33	Moderate
	158	LLSSPDPDV	118.238	217.90	Non-binder
SMM	546	KLLNNNLSL	3.61	114.54	Non-binder
	763	KLPDFSWEL	3.63	54.54	Moderate
	209	LLALKTLGV	3.79	61.33	Moderate
	726	YVYEVTKSI	4.92	75.63	Moderate
	130	SLCLANMSA	5.39	133.76	Non-binder
	521	GALDILEEV	5.5	219.83	Non-binder
NetMHC	763	KLPDFSWEL	23	54.54	Moderate
	440	IMHAIISPL	105	63.06	Moderate
	840	GLPAPEMYV	146	46.03	Moderate
	109	VMNSVIAQL	227	118.41	Non-binder
	545	KLLNNNLSL	249	114.54	Non-binder
	725	YVYEVTKSI	354	75.63	Moderate
MHCPred	836	GVIGGLPAP	16.75	97.92	Weak
	478	NSGGLEPLV	26.55	261.49	Non-binder
	209	LLALKTLGV	30.83	61.33	Moderate
	188	QELNAIPPI	38.64	84.37	Weak
	74	KLLTHEDKI	39.17	105.65	Weak
	546	KLLNNNLSL	42.66	114.54	Non-binder

The top six predicted peptides in each method and the corresponding PePSSI score and binding class are shown.

[a] PREDEP: energy score; SYFPEITHI: score; ProPred-I: real score; SMM: $\ln(IC_{50})$; NetMHC: affinity (nM); MHCPred: predicted IC_{50} value (nM).

[b] Strong binder = PePSSI score < 50; moderate binder = $50 \leq$ PePSSI score < 80; weak binder = $80 \leq$ PePSSI score < 110; non-binder = PePSSI score ≥ 110. Scores for strong and moderate binders are means of three PePSSI runs. All other scores are from a single PePSSI run.

Table 2
Comparison with PREDEP for the top 12 binding peptides predicted in PePSSI (listed in order from 1 to 12)

Sequence position	Peptide sequence	PePSSI score	PePSSI score (S.D.)	PREDEP rank
715	WCPPSDPDF	12.96	2.73	42
845	EMYVIDLMF	13.49	0.95	1
794	IFYHRALLF	14.35	3.29	2
9	VEPPPKDVF	14.86	2.17	591
196	ILDLLKSEY	16.47	3.25	396
567	IINDGFYDY	18.29	6.43	77
697	EEVMVVPKF	19.40	4.70	285
788	GHVKKGIFY	20.26	1.85	191
863	SREADLYRF	22.15	5.47	450
173	CIYNLVQDF	22.85	0.28	86
770	ELHISELKF	23.29	7.26	6
83	VRRNATMIF	23.64	4.21	140

and a standard error of 0.718. The error on the intercept is 0.522 and errors for the coefficients are both 0.006. The correlation is statistically weak, but the results below indicate that the equation provides a measure of the peptide affinity.

To place the binding score on an appropriate scale, the structures of a peptide known to be a strong HLA-A2 binder (GILGFVFTL; derived from the influenza matrix virus [35]) and a peptide known as a weak HLA-A2 binder (IISAVVGIL; the Gp2 peptide from HER2/Neu [12,13]) were predicted using PePSSI. Eqs. (1) and (2) were used to calculate a binding score for each peptide, using the PePSSI-predicted structures and averaging over six runs of PePSSI. A statistically significant difference was found between the strong and weak binder (binding scores: 37.7 ± 9.6 versus 86.3 ± 21.1; $n = 6$ for each peptide; independent t-test, $p = 0.0003$). From this analysis, a strong-binding peptide is defined as one with a binding score of <50 ($37.7 + 9.6$, rounded up to the nearest 10). Similarly, weak-binding peptides are defined as those with a PePSSI binding score of ≥ 80 and <110, based on the data for the Gp2 peptide. Peptides with predicted binding scores of ≥ 50 and <80 are considered to be moderate binders, and those with binding scores ≥ 110 are non-binders.

3.2. Overview of the PePSSI prediction for KU-CT-1L

The KU-CT-IL sequence has 872 amino acids [32], giving a total of 864 peptides with 9 amino acids. In the first PePSSI run, 77 peptides (8.9%) were predicted to be strong HLA-A2 binders (binding score < 50), 123 peptides (14.2%) were predicted to be moderate binders ($50 \leq$ binding score < 80), 159 peptides (18.4%) were predicted to be weak binders ($80 \leq$ binding score < 110), 500 peptides (57.9%) were predicted to be non-binders (although PePSSI was able to locate possible conformers in the HLA-A2-binding groove), and 5 peptides (0.6%) had no conformations that fitted the binding groove (all 5 peptides had a tryptophan residue in position 9; these were the only 5 peptides with this particular characteristic). The strong binders (77 peptides) were subjected to two further PePSSI runs. Of these peptides, 41 had predicted scores of <50 in runs 2 and 3, and 55 had a mean score of <50 averaged over runs 1–3. Of the 28 peptides (3.2% of the total peptide pool) with a score of <30 in run 1, none had a score of ≥ 50 in run 2 or 3, and all had a mean score of <40 over runs 1–3; hence, these peptides were convincingly predicted to be strong binders.

3.3. Algorithm comparison

The top six predicted peptides in each of the six algorithms are shown in Table 1, with the corresponding PePSSI score for each peptide. Broadly the algorithms are based on a motif-based approach (SYFPEITHI, ProPred-I), use of a training set of binding data (SMM, NetMHC), a QSAR approach (MHCPred), and a mixed structure/motif-based approach (PREDEP). Peptide 209 (LLALKTLGV) and peptide 763 (KLPDFSWEL) were each identified as probable HLA-A2 binders by three of the four algorithms (SYFPEITHI, ProPred-I, SMM and NetMHC) that broadly depend on experimental data (either directly, or indirectly through use of identified peptide motifs). Peptides 209 and 763 both contain the characteristic hydrophobic anchor residues at positions 2 and 9 that are associated with peptide binding to HLA-A2. PePSSI identifies both peptides as moderate binders (Table 1). Peptide 763 is also identified as a binder by the PREDEP algorithm. We return to the PePSSI structure prediction for peptide 763 below.

The striking feature of the comparison in Table 1 is the strong agreement between the PePSSI and PREDEP predictions (structure-based), and the relative lack of agreement between PePSSI or PREDEP with the other algorithms. Five of the top six PREDEP-predicted peptides are also predicted to be strong binders by PePSSI; in fact, four of the five are in the 3.2% of peptides identified by PePSSI as particularly strong binders (binding score <40 in all three runs), and the fifth (peptide 847, YVIDLMFHP) has an average PePSSI binding score of 40.25. The top 12 scoring peptides in the PePSSI prediction are shown in Table 2. Of particular note, the peptides ranked second and third in the PePSSI prediction (peptides 845 and 794, respectively) are ranked first and second, respectively, in the PREDEP prediction; it is also of note that neither of these peptides is predicted by motif-based approaches. There are some differences between the PePSSI and PREDEP predictions, but the overall agreement between the two structure-based algorithms is indicated by the

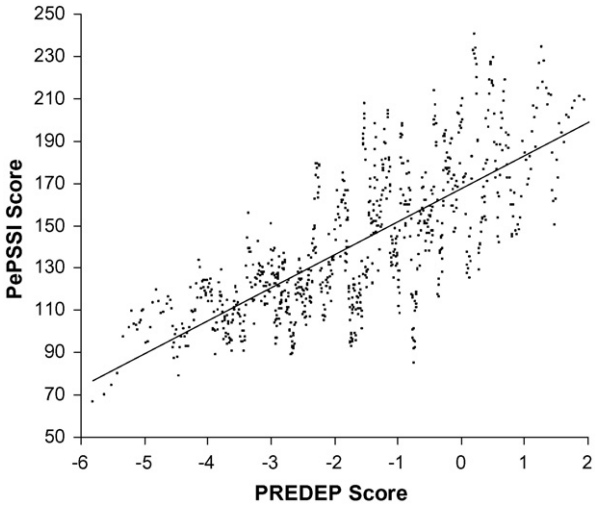

Fig. 1. Correlation of PePSSI and PREDEP scores for peptides derived from the KU-CT-1L sequence. The peptides were arranged in order based on PREDEP rankings and running averages in overlapping windows of 10 peptides were calculated for the PePSSI and PREDEP scores. The plot includes data for 859 of the 864 9-amino acid peptides comprising the KU-CT-1L sequence; five peptides (all with tryptophan at position 9) were excluded, since PePSSI found no allowable conformations of these peptides with HLA-A2. Linear regression analysis of the data gave a correlation coefficient of 0.546.

data comparison in Fig. 1. Using 859 of the 864 peptides (with exclusion of the 5 peptides for which PePSSI was unable to locate a binding conformation), the peptides were aligned in PREDEP rank order, and running averages over windows of 10 peptides were calculated for the PREDEP and PePSSI scores. Linear regression of these data (Fig. 1) gave a correlation coefficient of 0.546.

3.4. PePSSI-predicted peptide structures

The PePSSI algorithm gives output for 25 conformers of each peptide in a single run [30]. To group these conformers, we categorize peptide side chains as oriented away from the MHC binding groove (code 1: see below and Table 3), facing the α2 helix (code 2), facing the α1 helix (code 3), and facing the β-sheet (code 4) (see Fig. 2, Table 3 and Ref. [30]). This ''up–down-sideways'' approach is useful both for structural description and for understanding the key amino acids that mediate interactions with the MHC molecule and those (oriented away from the groove) that might make contacts with the T-cell receptor.

Table 3
Summary of PePSSI-predicted conformations of selected peptides

Sequence position	Peptide sequence	Conformation code[a]	Number of occurences[b]
763	KLPDFSWEL	132321234	39
845	EMYVIDLMF	132(1/3)2(1/3)2(1/3)4	40
794	IFYHRALLF	132(1/3)2(1/3/4)234	36
715	WCPPSDPDF	132(1/3)2(1/3)234[c]	24
		132134134[c]	33

[a] 1 = oriented away from the MHC binding groove; 2 = facing the α2 helix; 3 = facing the α1 helix; 4 = facing the β-sheet (oriented down into the groove). Each number represents the side chain orientation at the particular position in the peptide. Numbers in parentheses indicate multiple orientations at the corresponding position.

[b] Number of occurrences of the designated conformer out of 75 predicted conformers for each peptide (25 in each of 3 PePSSI runs).

[c] Two conformers were predicted for this peptide.

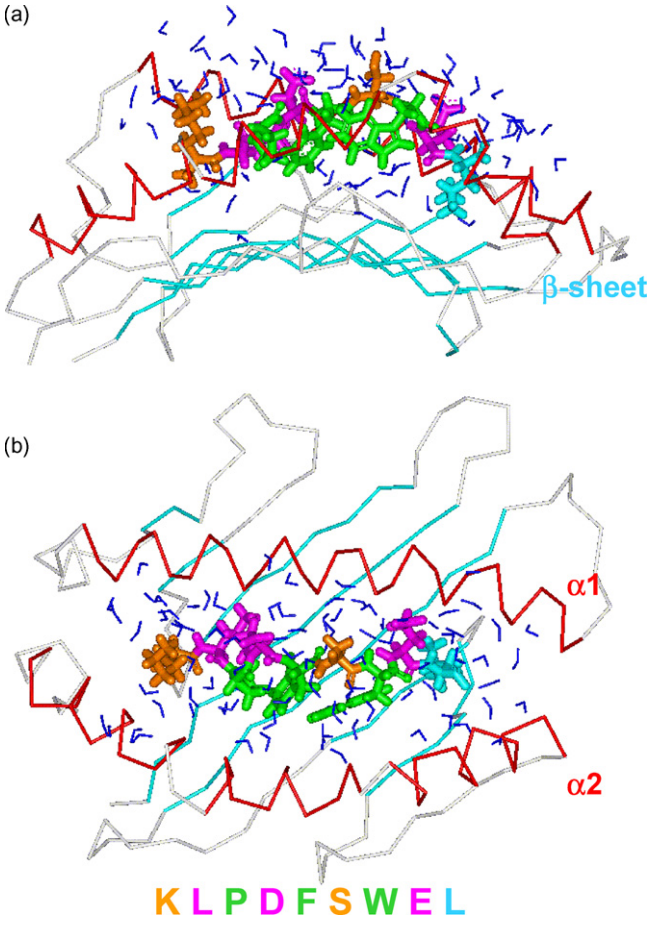

Fig. 2. Predicted structure for peptide 763 (KLPDFSWEL) bound to HLA-A2, viewed (a) from the side of the complex, with the α2 helix of the MHC molecule in the foreground and (b) looking into the binding groove. The α-helices (shown in red) of the α1 and α2 domains of the MHC molecule form the sides of the binding groove and the β-sheet (light blue) forms the floor of the groove. Water molecules are shown in blue. The amino acids of the peptide are colored based on their orientation with respect to the binding groove: orange, facing away from the groove (code 1 in Table 3); green, facing the α2 helix (code 2); pink, facing the α1 helix (code 3); light blue, oriented towards the β-sheet (code 4). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

Using this approach, we examined four peptides from the PePSSI analysis of the KU-CT-1L sequence: peptide 763 (KLPDFSWEL) as an example of a sequence with ''classical'' HLA-A2 anchor residues at positions 2 and 9; and the top three scoring peptides from PePSSI (Table 2). Since three PePSSI runs were recorded for each peptide, 75 structures were used in the analysis. For three of the peptides, a dominant conformation was adopted by about half the predicted peptide conformers, but for peptide 715 (WCPPSDPDF) two discrete conformer clusters were obtained (Table 3). In the following sections, we discuss further details of the predicted structures.

3.5. Peptide 763 (KLPDFSWEL)

This peptide has a classical sequence motif for binding to HLA-A2, as described above. As shown in Fig. 2, the N-terminal lysine side chain is oriented away from the groove and the C-terminal leucine side chain is oriented into the groove (towards the β-sheet). Of the non-terminal hydrophobic residues, P3, F5 and W7 are oriented towards the α2 helix and L2 is oriented towards the α1 helix (Fig. 2); these locations are consistent with those in the X-ray structure of the HLA-A2 complex of the influenza matrix peptide GILGFVFTL [36], which has a strong affinity for HLA-A2. The side chains of the three non-terminal hydrophilic residues, D4, S6 and E8 are oriented into the solvent (away from the groove); although D4 and E8 are formally categorized as pointing towards the α1 helix, it is evident from Fig. 2(a) that the carboxylate groups of these side chains are both solvent exposed.

3.6. Peptide 845 (EMYVIDLMF) and peptide 794 (IFYHRALLF)

These peptides were predicted to be strong HLA-A2 binders by PePSSI and PREDEP (Table 2), but were not identified as such by any of the other algorithms. This is of interest, since these peptides do not contain the typical anchor residues at positions 2 and 9 for HLA-A2-binding peptides (although both peptides have hydrophobic residues at these positions). The predicted peptide structures (as bound to HLA-A2) are shown in Fig. 3, using the same views as those for peptide 763 in Fig. 2, but with the HLA-A2 protein and water removed for clarity. It is apparent that peptides 845 and 794 both adopt a similar conformation to peptide 763; the amino acids at positions 1, 2, 3, 5, 7, 8 and 9 adopt similar orientations in all three peptides, and these provide the basis for the strong binding to HLA-A2. For peptide 794 (IFYHRALLF), the methylene groups of the R5 side chain are able to fulfill the hydrophobic requirement at this position, with solvent exposure of the guanidinium group. The amino acids at positions 4 and 6 differ between peptides 845 and 794: a hydrophobic valine at position 4 and a hydrophilic (negatively charged) aspartic acid at position 6 in peptide 845, compared to a histidine at position 4 and an alanine at position 6 in peptide 794. The PePSSI algorithm is able to recognize this difference through prediction of a different conformation for the central region of the two peptides; in each conformation, the hydrophilic residue is oriented out of the binding groove (central orange-colored residues in Fig. 3(a) and (c), respectively) and the hydrophobic residue is oriented towards the α1 helix.

3.7. Peptides 715 (WCPPSDPDF)

This peptide is predicted to be a strong binder by PePSSI, but not by any other algorithm, including PREDEP. The PePSSI prediction indicated two discrete conformers for peptide 715 (Table 3; Fig. 4). Conformer 1 has similar side chain orientations to peptide 845, with the aspartic acid at position 6 oriented directly into the solvent (away from the binding groove) (Fig. 4(a)). Conformer 2 has a very different conformation, with P4 and P7 oriented away from the binding groove and D6 oriented towards the β-sheet (Fig. 4(c)). At first sight this conformation appears unlikely, given the solvent exposure of two proline side chains and the apparently

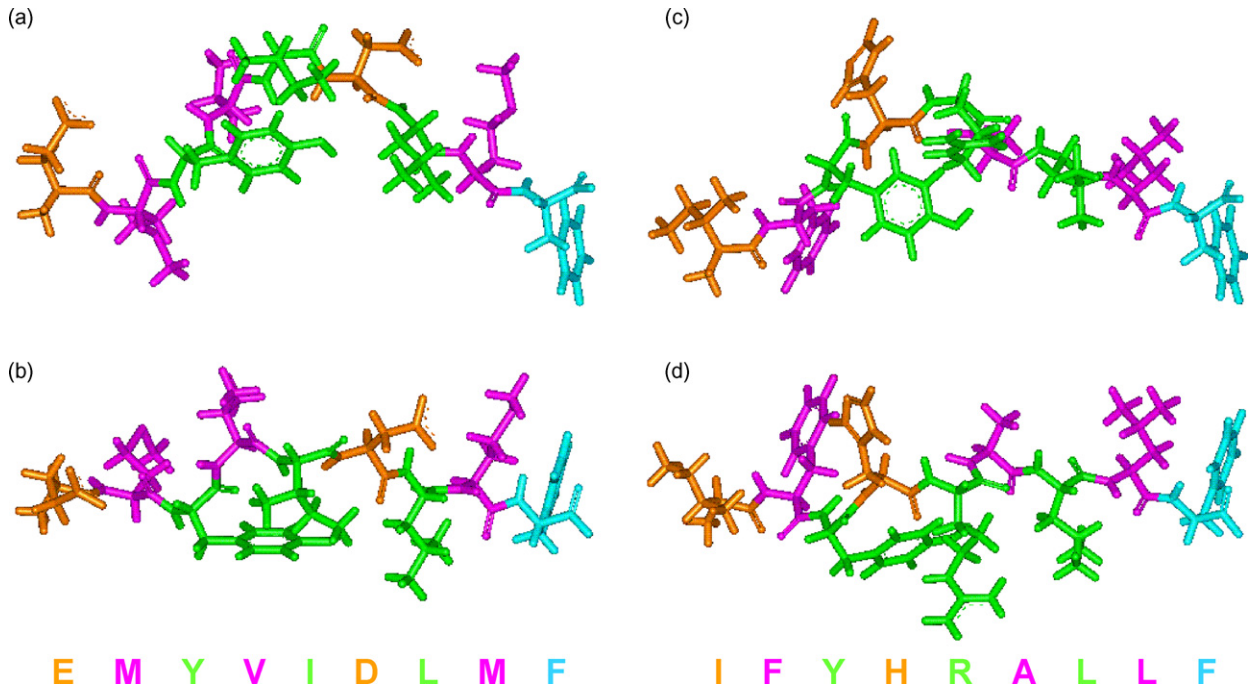

Fig. 3. Predicted structures for peptide 845 (EMYVIDLMF) and peptide 794 (IFYHRALLF) when bound to HLA-A2, viewed from the side of the peptide–MHC complex [(a) and (c), respectively] and looking into the binding groove [(b) and (d), respectively]. The coloring of amino acids of the peptide is similar to that described in the legend to Fig. 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

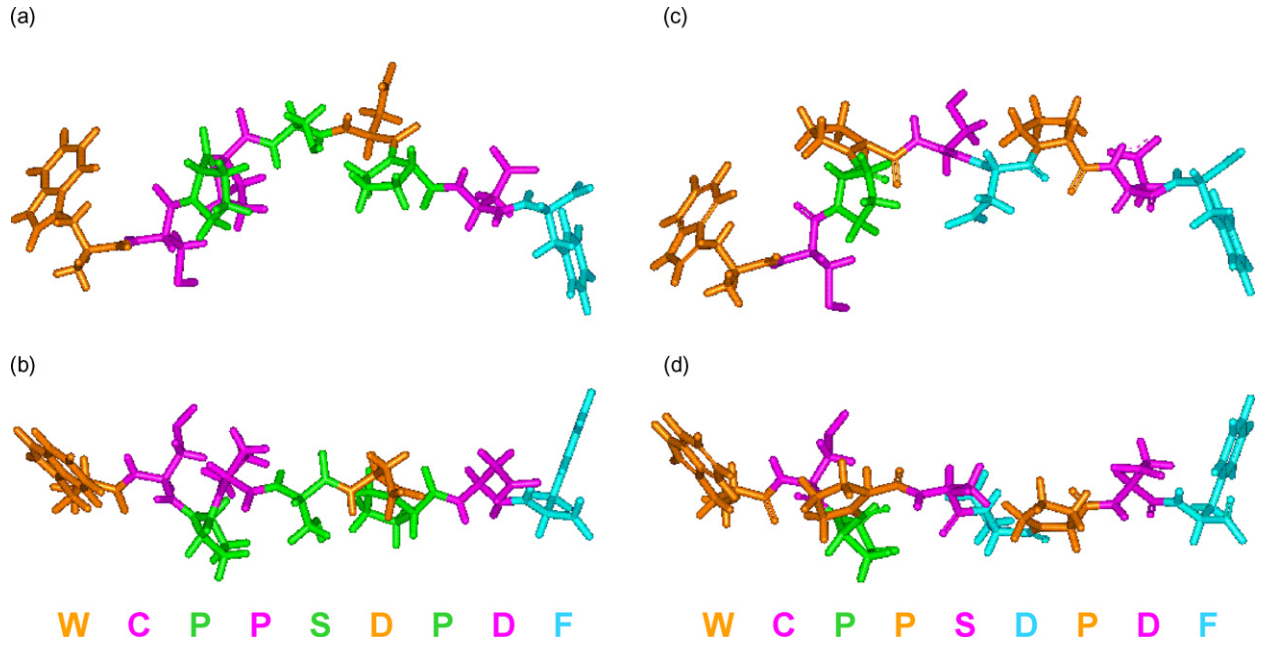

Fig. 4. Predicted structures for conformers 1 and 2 (Table 3) of peptide 715 (WCPPSDPDF) when bound to HLA-A2, viewed from the side of the peptide–MHC complex [(a) and (c), respectively] and looking into the binding groove [(b) and (d), respectively]. The coloring of amino acids of the peptide is similar to that described in the legend to Fig. 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

unfavorable location for D6, but it is of note that the HLA-A2-binding groove contains a centrally located arginine residue (R97), and in conformer 2 there is a strong predicted electrostatic interaction between R97 and the peptide D6 side chain.

4. Discussion

The results indicate good agreement between the PePSSI and PREDEP algorithms, both of which incorporate structure into the prediction of peptide binding affinity for class I MHC molecules. Experimental validation of these results is clearly required, but we believe that the agreement between PePSSI and PREDEP is an interesting outcome, since the algorithms were developed independently and use different approaches for structure prediction. PREDEP incorporates a threading approach based on a fixed peptide backbone, in addition to supplementary motif information [22]. In PePSSI, we introduced considerably more flexibility (at a computational cost; each peptide run requires about 1 h of running time on a standard desktop PC); hence, the peptide conformation is almost completely flexible and some variation of the side chain conformations of key MHC amino acids is also allowed [30].

Comparisons of sequence-based methods for prediction of MHC-binding peptides have shown that no single method is superior [24,37–40]. These findings have been clarified in a recent study by Peters et al. [41], in which a comparison of most publicly available MHC–peptide prediction algorithms indicated that different algorithms have greatest accuracy for different MHC molecules (HLA alleles) and peptide lengths. Sequence-based approaches allow rapid scanning of a protein sequence and identification of potential binding peptides that have similar motifs to peptides known to bind to a given HLA allele. Such methods are considerably faster than a structure-based approach, but they include the assumption that a motif will behave similarly in any peptide sequence context, which may not always be correct. Nonetheless, these approaches have been successful in identifying peptide epitopes, and a motif-based approach with subsequent structural prediction of identified peptides using PePSSI may be a viable strategy. However, a structural approach may be required for detection of binding peptides with sequences inconsistent with known binders, as also suggested by Altuvia and Margalit [22].

The PePSSI results for peptides from KU-CT-1L indicate a relatively conserved conformation for all the strong-binding peptides, and so the relatively less flexible PREDEP approach may be sufficient to obtain a good estimate of the binding conformation. However, PePSSI may allow assessment of subtle differences in binding conformations (for example, between the central regions of peptides 845 and 794), and this may be of importance for subsequent prediction of side chains that might interact with a T-cell receptor (TCR), which is of particular importance in vaccine design [11]. The two conformations predicted for peptide 715 are also of interest. Conformation 1 has the characteristic features of other HLA-A2 binders, whereas conformation 2 has a very different, but still apparently favorable, conformation. It is possible that a switch from conformation 1 to 2 on TCR binding with orientation of the proline residues of peptide 715 towards the T-cell receptor may offer a hydrophobic surface that provides an improved interaction with the T-cell receptor, while desolvation of D6 is favored through adoption of an electrostatic interaction with R97 of the MHC molecule. This is speculative, but peptide 715 may be of interest for experimental testing.

Finally, we note that the PePSSI scoring system requires more development, and this is a goal for the next version of the

algorithm. The relationship of the affinity for HLA-A2 with the number of contacts is supported by the general hydrophobicity of peptides that bind to this particular MHC molecule. However, it seems likely that hydrogen bonding and electrostatic association should be included in the binding equation, but we have been unable to find a significant relationship of affinity with these interactions (data not shown). Similarly, the relationship of affinity with a reduced number of peptide–MHC bridging water molecules is consistent with favorable binding of hydrophobic peptides, and further analysis of the enthalpic and entropic components associated with water-molecule trapping at the peptide–MHC interface [42] is likely to produce a more sophisticated scoring function.

Acknowledgement

AJS was supported by a fellowship from the American Foundation for Pharmaceutical Education.

References

[1] M. Schuster, A. Nechansky, R. Kircheis, Cancer immunotherapy, Biotechnol. J. 1 (2006) 138–147.
[2] C.H. Huber, T. Wolfel, Immunotherapy of cancer: from vision to standard clinical practice, J. Cancer Res. Clin. Oncol. 130 (2004) 367–374.
[3] K. Evel-Kabler, S.Y. Chen, Dendritic cell-based tumor vaccines and antigen presentation attenuators, Mol. Ther. 13 (2006) 850–858.
[4] S.K. Tey, C.M. Bollard, H.E. Heslop, Adoptive T-cell transfer in cancer immunotherapy, Immunol. Cell Biol. 84 (2006) 281–289.
[5] S.A. Rosenberg, R.M. Sherry, K.E. Morton, J.C. Yang, S.L. Topalian, R.E. Royal, U.S. Kammula, N.P. Restifo, M.S. Hughes, S.L. Schwarz, L.T. Ngo, S.A. Mavroukakis, D.E. White, Altered CD8(+) T-cell responses when immunizing with multiepitope peptide vaccines, J. Immunother. 29 (2006) 224–231.
[6] P. Romero, J.-C. Cerottini, D.E. Speiser, Monitoring tumor antigen specific T-cell responses in cancer patients and phase I clinical trials of peptide-based vaccination, Cancer Immunol. Immunother. 53 (2004) 249–255.
[7] A.K. Abbas, A.H. Lichtman, Cellular and Molecular Immunology, Elsevier Sauders, Philadelphia, PA, 2005.
[8] M.G. Rudolph, R.L. Stanfield, I.A. Wilson, How TCRs bind MHCs, peptides, and coreceptors, Annu. Rev. Immunol. 24 (2006) 419–466.
[9] D.R. Madden, The three-dimensional structure of peptide–MHC complexes, Annu. Rev. Immunol. 13 (1995) 587–622.
[10] M.P. Davenport, K.J. Smith, D. Barouch, S.W. Reid, W.M. Bodnar, A.C. Willis, D.F. Hunt, A.V. Hill, HLA class I binding motifs derived from random peptide libraries differ at the COOH terminus from those of eluted peptides, J. Exp. Med. 185 (1997) 367–371.
[11] W.S. Meng, L.H. Butterfield, Rational design of peptide-based tumor vaccines, Pharm. Res. 19 (2002) 926–932.
[12] J.J. Kuhns, M.A. Batalia, S. Yan, E.J. Collins, Poor binding of a HER-2/neu epitope (GP2) to HLA-A2.1 is due to a lack of interactions with the center of the peptide, J. Biol. Chem. 274 (1999) 36422–36427.
[13] Y. Tanaka, K.D. Amos, H.G. Joo, T.J. Eberlein, P.S. Goedegebuure, Modification of the HER2/NEU-derived tumor antigen GP2 improves induction of GP2-reactive cytotoxic T lymphocytes, Int. J. Cancer 94 (2001) 540–544.
[14] H. Rammensee, J. Bachmann, N.P. Emmerich, O.A. Bachor, S. Stevanovic, SYFPEITHI: database for MHC ligands and peptide motifs, Immunogenetics 50 (1999) 213–219.
[15] K.C. Parker, M.A. Bednarek, L.K. Hull, U. Utz, B. Cunningham, H.J. Zweerink, W.E. Biddison, J.E. Coligan, Sequence motifs important for peptide binding to the human MHC class I molecule, HLA-A2, J. Immunol. 149 (1992) 3580–3587.
[16] V. Brusic, G. Rudy, L.C. Harrison, Prediction of MHC binding peptides using artificial neural networks, in: Complex Systems: Mechanism of Adaptation, IOS Press, Amsterdam, 1994, 253–260.
[17] H.P. Adams, J.A. Koziol, Prediction of binding to MHC class I molecules, J. Immunol. Methods 185 (1995) 181–190.
[18] M. Milik, D. Sauer, A.P. Brunmark, L. Yuan, A. Vitiello, M.R. Jackson, P.A. Peterson, J. Skolnick, C.A. Glass, Application of an artificial neural network to predict specific class I MHC binding peptide sequences, Nat. Biotechnol. 16 (1998) 753–756.
[19] I.A. Doytchinova, D.R. Flower, Towards the quantitative prediction of T-cell epitopes: CoMFA and CoMSIA studies of peptides with affinity for the class I MHC molecule HLA-A*0201, J. Med. Chem. 44 (2001) 3572–3581.
[20] O. Schueler-Furman, Y. Altuvia, A. Sette, H. Margalit, Structure-based prediction of binding peptides to MHC class I molecules: application to a broad range of MHC alleles, Protein Sci. 9 (2000) 1838–1846.
[21] J. Desmet, G. Meersseman, N. Boutonnet, J. Pletinckx, K. De Clercq, M. Debulpaep, T. Braeckman, I. Lasters, Anchor profiles of HLA-specific peptides: analysis by a novel affinity scoring method and experimental validation, Proteins 58 (2005) 53–69.
[22] Y. Altuvia, H. Margalit, A structure-based approach for prediction of MHC-binding peptides, Methods 34 (2004) 454–459.
[23] H. Singh, G.P.S. Raghava, ProPred1: prediction of promiscuous MHC class-I binding sites, Bioinformatics 19 (2003) 1009–1014.
[24] B. Peters, W. Tong, J. Sidney, A. Sette, Z. Weng, Examining the independent binding assumption for binding of peptide epitopes to MHC-I molecules, Bioinformatics 19 (2003) 1765–1772.
[25] S. Buus, S.L. Lauemoller, P. Worning, C. Kesmir, T. Frimurer, S. Corbet, A. Fomsgaard, J. Hilden, A. Holm, S. Brunak, Sensitive quantitative predictions of peptide–MHC binding by a 'Query by Committee' artificial neural network approach, Tissue Antigens 62 (2003) 378–384.
[26] P. Guan, I.A. Doytchinova, C. Zygouri, D.R. Flower, MHCPred: bringing a quantitative dimension to the online prediction of MHC binding, Appl. Bioinform. 2 (2003) 63–66.
[27] K.J. Smith, S.W. Reid, K. Harlos, A.J. McMichael, D.I. Stuart, J.I. Bell, E.Y. Jones, Bound water structure and polymorphic amino acids act together to allow the binding of different peptides to MHC class I HLA-B53, Immunity 4 (1996) 215–228.
[28] W.S. Meng, H. von Grafenstein, I.S. Haworth, A model of water structure inside the HLA-A2 peptide binding groove, Int. Immunol. 9 (1997) 1339–1346.
[29] W.S. Meng, H. von Grafenstein, I.S. Haworth, Water dynamics at the binding interface of four different HLA-A2–peptide complexes, Int. Immunol. 12 (2000) 949–957.
[30] H.-H. Bui, A.J. Schiewe, H. von Grafenstein, I.S. Haworth, Structural prediction of peptides binding to MHC class I molecules, Proteins 63 (2006) 43–52.
[31] H.-H. Bui, A.J. Schiewe, I.S. Haworth, WATGEN: an algorithm for modeling water networks at protein–protein binding interfaces. J. Comp. Chem., in press.
[32] T. Okada, M. Akada, T. Fujita, T. Iwata, Y. Goto, K. Kido, T. Okada, Y. Matsuzaki, K. Kobayashi, S. Matsuno, M. Sunamura, Y. Kawakami, A novel cancer testis antigen that is frequently expressed in pancreatic, lung, and endometrial cancers, Clin. Cancer Res. 12 (2006) 191–197.
[33] C.E. Clark, R.H. Vonderheide, Cancer-testis antigens in tumor biology and immunotherapy, Cancer Biol. Ther. 5 (2006) 1226–1227.
[34] I.A. Doytchinova, D.R. Flower, Physicochemical explanation of peptide binding to HLA-A*0201 major histocompatibility complex: a three-dimensional quantitative structure–activity relationship study, Proteins 48 (2002) 505–518.
[35] M.A. Bednarek, S.Y. Sauma, M.C. Gammon, G. Porter, S. Tamhankar, A.R. Williamson, H.J. Zweerink, The minimum peptide epitope from the influenza virus matrix protein, J. Immunol. 127 (1991) 4047–4053.
[36] D.R. Madden, D.N. Garboczi, D.C. Wiley, The antigenic identity of peptide–MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2, Cell 75 (1993) 693–708.
[37] K. Gulukota, J. Sidney, A. Sette, C. DeLisi, Two complementary methods for predicting peptides binding major histocompatibility complex molecules, J. Mol. Biol. 267 (1997) 1258–1267.

[38] M. Bhasin, G.P. Raghava, Prediction of CTL epitopes using QM, SVM and ANN techniques, Vaccine 22 (2004) 3195–3204.
[39] R.R. Mallios, A consensus strategy for combining HLA-DR binding algorithms, Hum. Immunol. 64 (2003) 852–856.
[40] K. Yu, N. Petrovsky, C. Schonbach, J.Y. Koh, V. Brusic, Methods for prediction of peptide binding to MHC molecules: a comparative study, Mol. Med. 8 (2002) 137–148.
[41] B. Peters, H.-H. Bui, S. Franklid, M. Nielsen, C. Lundegaard, E. Kostem, D. Basch, L. Kasper, M. Harndahl, W. Fleri, S.S. Wilson, J. Sidney, O. Lund, S. Buus, A. Sette, A community resource benchmarking prediction of peptide binding to MHC-I molecules, PLoS Comp. Biol. 2 (2006) e65.
[42] P.M. Petrone, A.E. Garcia, MHC–peptide binding is assisted by bound water molecules, J. Mol. Biol. 338 (2004) 419–435.