

Review Article

The pockets guide to HLA class I molecules

Andrea T. Nguyen^{1,2}, Christopher Szeto^{1,2} and  Stephanie Gras^{1,2}

¹Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Bundoora, Victoria 3083, Australia; ²Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia

Correspondence: Stephanie Gras (s.gras@latrobe.edu.au) or Christopher Szeto (c.szeto@latrobe.edu.au)



Human leukocyte antigens (HLA) are cell-surface proteins that present peptides to T cells. These peptides are bound within the peptide binding cleft of HLA, and together as a complex, are recognised by T cells using their specialised T cell receptors. Within the cleft, the peptide residue side chains bind into distinct pockets. These pockets ultimately determine the specificity of peptide binding. As HLAs are the most polymorphic molecules in humans, amino acid variants in each binding pocket influences the peptide repertoire that can be presented on the cell surface. Here, we review each of the 6 HLA binding pockets of HLA class I (HLA-I) molecules. The binding specificity of pockets B and F are strong determinants of peptide binding and have been used to classify HLA into supertypes, a useful tool to predict peptide binding to a given HLA. Over the years, peptide binding prediction has also become more reliable by using binding affinity and mass spectrometry data. Crystal structures of peptide-bound HLA molecules provide a means to interrogate the interactions between binding pockets and peptide residue side chains. We find that most of the bound peptides from these structures conform to binding motifs determined from prediction software and examine outliers to learn how these HLAs are stabilised from a structural perspective.

Introduction

The major histocompatibility complex (MHC) class I, known as human leukocyte antigen (HLA) class I in humans, is an essential surface molecule composed of a heavy chain and an invariant beta-2-microglobulin ($\beta 2m$) that presents peptides to T cells [1]. HLA molecules are critical for the selection and activation of T cells and play a key role in the immune response to many pathogens.

There are two main classes of HLAs: HLA class I (HLA-I) and class II (HLA-II) with major structural differences between the two. Most prominently is that the HLA-I peptide binding cleft is closed at the N and C termini, and therefore, restricts the bound peptide to an optimal length of 8–10 amino acids (reviewed in [2]). Both ends of the binding cleft of HLA-II is open-ended, and thus, has a preference for peptides that are >13 amino acids in length [3].

The peptides presented by HLA molecules can be derived from host proteins (self-peptides) or from a pathogenic source such as viruses and bacteria. HLA is the most polymorphic molecule in humans, with >22 000 HLA alleles reported to date [4]. This allows HLAs to present a wide range of pathogenic-derived peptides to T cells (reviewed in [5]) and can also help to limit any pathogenic mutant escape from the immune system at a population level.

Due to this extreme diversity, each individual will present a different set of peptides, termed peptide repertoire, to their T cells, and this has greatly limited the use of T cell-based vaccines or therapeutics, as they are not applicable to a wide population. Despite this limitation, research continues to provide a better understanding of the relationship between pathogens, HLA presentation and peptide binding, which can dictate the T cell response. As CD8⁺ T cells are responsible for recognising peptide-bound HLA-I, killing infected cells and thus clearing infection, we have focused our review on HLA-I molecules.

Received: 7 August 2021
Revised: 3 September 2021
Accepted: 6 September 2021

Version of Record published:
28 September 2021

The motivation behind characterising and understanding HLA binding pockets and their polymorphisms is to determine which peptides can bind to and be presented to the immune system. Interestingly, the polymorphisms that define each HLA variant have been found to impact peptide binding the most [6].

Although, there are many HLA alleles in existence, some alleles are more common in the population than others (e.g. HLA-A*02:01, HLA-A*24:02, HLA-B*35:01, etc.) and are often the subject of study for T cell based therapeutics. These HLA molecules have specificity for certain amino acid side chains in distinct binding pockets within the HLA binding cleft, which limits peptide sequences it can present to the immune system from any one pathogen.

Within the past two decades, over >600 structures of peptide-HLA (pHLA) complexes have been solved by X-ray crystallography and submitted to the Protein Data Bank (PDB). This allows insight into the binding patterns of peptides to specific HLA allotypes [7]. Interestingly, this still represents a very narrow slice of the diversity of HLA molecules (reviewed in [8]) but has been informative in understanding the rules of peptide binding specificity within specific HLA molecules.

HLA pockets description

HLA-I molecules possess six distinct binding pockets within the binding cleft termed as pockets A–F [9] (Figure 1), which allow peptide residue side chains to anchor or bind deeply within each pocket. Each HLA pocket has allotype-specific biochemical properties [10] based on the polymorphisms of each HLA allotype. The amino acid make-up of each binding pocket thus determines the peptide side chain specificity that can bind. Importantly, pockets B and F house the primary anchor residues of the peptide. These anchor residues form the main interactions between peptide and HLA, and is suggested to play a key role in stabilising the peptide-HLA complex [1,10–12]. In HLA-I, pocket B houses position 2 (P2) of the peptide, whilst pocket F usually accommodates the C-terminal residue of the peptide (P Ω) (Figure 1B). Here, we describe each of the six pockets that form the HLA binding cleft (Table 1), based on the structures of the HLA molecules obtained by X-ray crystallography. The A pocket is composed of 9 residues that traditionally helps bind the N-terminal group of the peptide, usually the first residue or P1, as well as forming the closed end or ‘wall’ of the N-terminal part of the cleft (Figure 1C). Large aromatic residues are often present at position 59, 167, 171 (for e.g. in HLA-A*02:01) that helps close off this end of the cleft, which is a notable structural difference relative to HLA-II that has an open-ended conformation in their cleft. The closed conformation constrains the peptide length that is often shorter in HLA-I than HLA-II molecules, with some exceptions [2]. The B pocket consists of 10 residues (Table 1) with residues 9, 45, 63, 66, 67, 70, and 99 being the key residues that determine side chain specificity. The B pocket is binding one of the two primary anchor residues of the peptide, at P2, and determines the nature of the P2 residue able to bind to each HLA allotype (Figure 1D). The C and D pockets face each other in the middle of the cleft and are composed of 5 and 6 residues, respectively (Table 1). The C pocket is located against the α 1-helix and the D pocket against the α 2-helix of the HLA, both pockets bind secondary anchor residues at P3 and/or P5/6 depending on the peptide (Figure 1E). Some HLAs favour peptides with secondary anchor residues, for e.g. HLA-B*08:01 with a P5-R forming a salt bridge with Asp7 and Asp9 from the cleft (Figure 2A) [13]. The E pocket is formed by 5 residues and also partially binds the P5/6 secondary anchor residue as well as C-terminal residues of the peptide. The F pocket is composed of residues 77, 80, 81, 84, 95, 116, 123, 143, 146, and 147 (Figure 1F). Residues 77, 80, 81 and 116 are considered the key residues that determine the specificity of P Ω [10,12]. In addition, and as for pocket A, the last pocket also closes the C-terminal end of the cleft, with large aromatic residues at position 80, 81, 84, 123, 143 (as observed in HLA-A*02:01). The Tyr84 is conserved in about a third of HLA-I molecules [4], and replaced by Phe84 in ~10% of HLA-I, and can be used as a switch that opens to enable the binding of longer peptides [14,15]. Interestingly, a large bulky residue at position 84 is also observed in lipid and metabolite antigen binding MHC-like molecules CD1 and MR1, respectively. The Tyr84 is shared in human CD1a, while CD1b has a Phe84, CD1c has a His84 [16], and MR1 has also a His84 at the end of the C-terminal part of the antigen binding cleft.

Although it has been determined that the main anchor residues are necessary for binding, secondary anchors can also contribute significantly to overall binding as well (reviewed in [17,18]) and improve the overall stability of the peptide-HLA complex (pHLA).

HLA supertypes classification

The structural features of HLA pockets brought about the concept of HLA supertypes, introduced in the 90’s [19,20], which classified a number of HLA-A and HLA-B allotypes into 9 HLA-I supertypes [10] (Table 2).

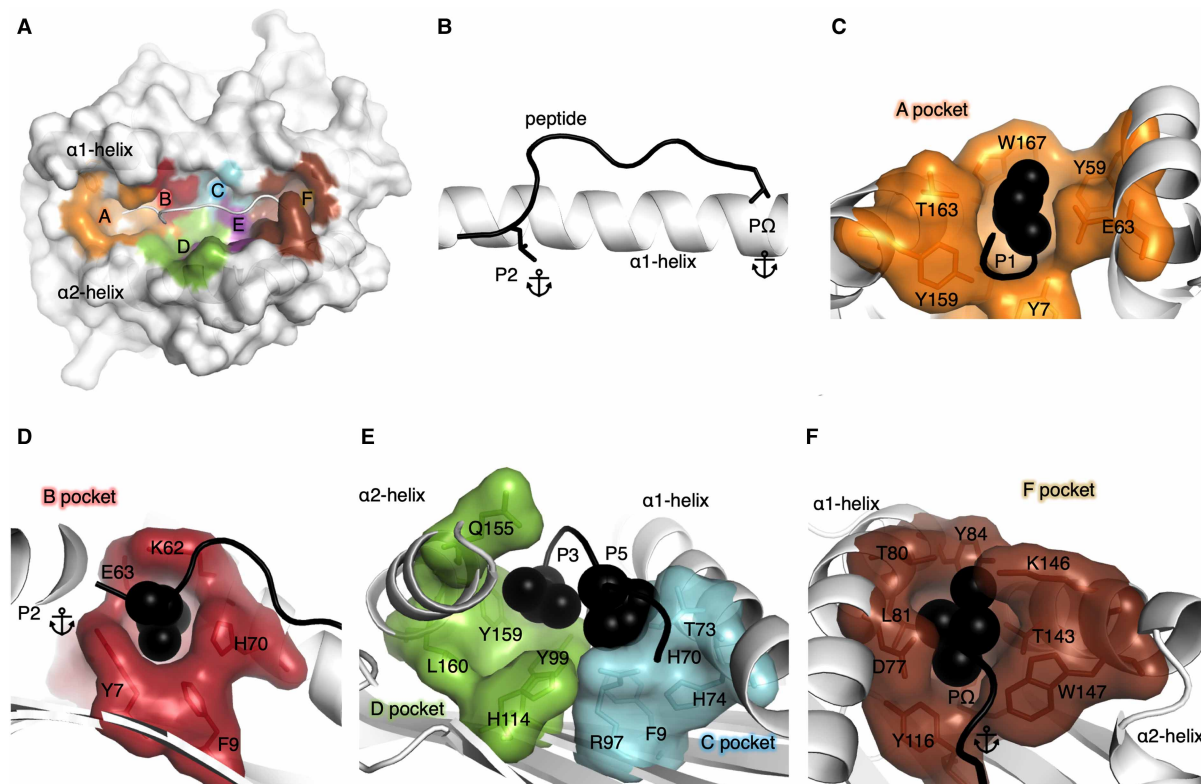


Figure 1. HLA pockets representation.

(A) Surface representation of a HLA molecule (white) with the pockets within the cleft shown in different colours. Pocket A (orange), B (red), C (cyan), D (green), E (purple), F (brown) and the peptide shown as a cartoon in white. (B) Peptide represented as a black cartoon with primary anchor residues P2 and P Ω shown as stick against the α 1-helix of the HLA shown as a white cartoon. (C) The A pocket of the HLA represented by an orange surface with the residues making up the pocket shown as orange sticks and the surrounding residues represented by a white cartoon. The P1 residue of the peptide is represented by black spheres nested into the A pocket. (D) The B pocket of the HLA represented by a red surface with the residues making up the pocket shown as red sticks and the surrounding residues represented by a white cartoon. The primary anchor P2 residue of the peptide is represented by black spheres. (E) The C pocket of the HLA represented by a cyan surface and residues as sticks, while the D pocket is represented by a green surface with the residues making up the D pocket shown as green sticks. The P3 and P5 of the peptide is represented by black spheres. (F) The F pocket of the HLA represented by a brown surface with the residues making up the pocket shown as brown sticks and the surrounding residues represented by a white cartoon. The P Ω of the peptide is represented by black spheres.

Table 1 Residues forming the HLA class I molecule pockets

Pocket	Residues	Role of the pocket
A	5, 7, 59, 63, 66, 159, 163, 167, 171	Wall of the N-terminal part of the binding cleft, bind P1 residue
B	7, 9, 24, 34, 45, 63, 66, 67, 70, 99	Bind primary anchor residue P2
C	9, 70, 73, 74, 97	Bind secondary anchor residue at P3 and P5/P6 when presents, face pocket D
D	99, 114, 155, 156, 159, 160	Bind secondary anchor residue at P3 and P5/P6 when presents, face pocket C
E	97, 114, 147, 152, 156	Overlap with C/D pockets and contact secondary anchor residue at P5/P6 when presents and the C-terminal part of the peptide
F	77, 80, 81, 84, 95, 123, 143, 146, 147	Bind primary anchor residue P Ω , wall of the C-terminal part of the binding cleft

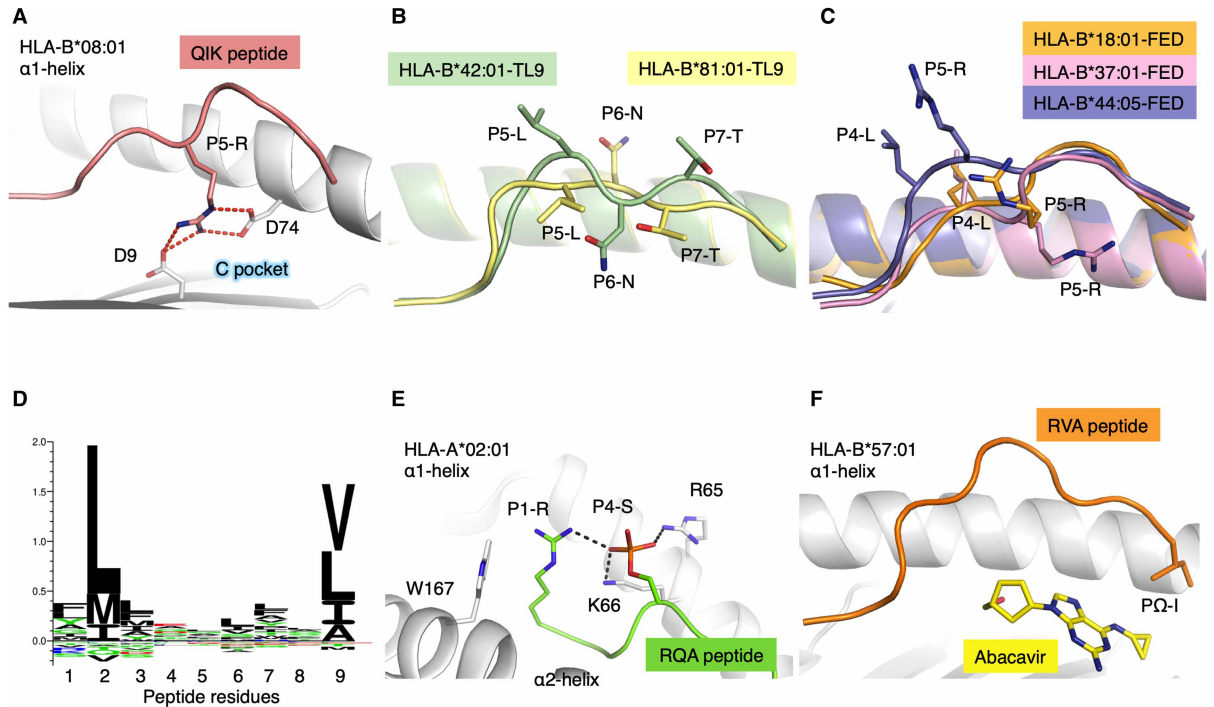


Figure 2. HLA polymorphism impacts on peptide presentation and binding.

(A) The α 1-helix of HLA-B*08:01 shown as a white cartoon with the residues interacting with the QIK peptide (salmon cartoon) shown as stick. The P5 of the QIK peptide is shown as stick, forming salt bridges (red dashed lines) with the Asp9 and Asp74 of the HLA molecule. (B) Overlay of HLA-B*42:01 (green cartoon) and HLA-B*81:01 (yellow cartoon) each presenting the TL9 peptide, in a distinct conformation, shown as cartoon (green and yellow, respectively) with P5, P6 and P7 of the peptide shown as stick. (C) Overlay of HLA-B*18:01 (orange cartoon), HLA-B*37:01 (pink cartoon) and HLA-B*44:05 (purple cartoon) presenting the FED peptide shown as cartoon (orange, pink and purple, respectively) with P4 and P5 of each peptide shown as stick. (D) Distribution of preferred peptide residues located into the B and F pockets of HLA-A*02:01 using Seq2logo2.0 [69]. (E) HLA-A*02:01 (white cartoon) with select residues shown as stick presenting the RQA peptide (green cartoon) with P1 shown as stick and P4 represented as an orange stick. The P4-S phosphorylated is represented as stick forming hydrogen bonds (black dashed lines) with the R65 from the HLA and the P1-R from the peptide. (F) HLA-B*57:01 (white cartoon) presenting the RVA peptide (orange cartoon) with the P Ω shown as stick, in the presence of Abacavir in the F pocket of the HLA (yellow stick).

They analysed the key residues that made up the B and F pockets to understand the type of anchor residue that the pockets preferred and used this to group HLAs into clusters of supertypes. This provided an index for some of the first peptide and epitope-based approaches for vaccine development [21–23].

Theoretically, peptides that bind to one specific HLA within a supertype, would also likely bind to other HLAs in the same supertype due to sequence similarity of the HLA cleft. However, in practice, the classification is not always clear-cut, with many HLAs in the same supertype binding promiscuously to peptides of different supertypes [24–27]. Additionally, the use of this method requires each HLA to be classified within a HLA supertype. The work laid out by Sette & Sidney was further expanded to classify 750 HLA's in total, but by 2007 the number of HLA-I's that had been discovered had reached ~1500 and now >22 000 [4]. This later rendition incorporated binding data into the HLA supertype classification and reshuffled some HLAs into different supertypes. Many different approaches used by several groups have yielded some new supertypes (for e.g. A6X supertype) or reshuffling of allotypes within existing supertype classifications [28–33] (Table 2).

Today, there are many clinical and immunological studies that use HLA supertypes as correlates or predictors of disease protection/severity [34–37], with current classifications incorporating up to 15 supertypes (Table 2). This highlights the simplicity and ease of use of supertypes to draw a relationship between HLA

peptide restriction and disease outcome. In the context of peptide-based vaccines, peptide presentation targeted to a specific HLA supertype can give an indication of the potential coverage within a given population; however, we should be mindful that supertype classifications does not guarantee peptide binding in all their HLA denominations. In addition, if one peptide is binding to multiple HLA molecules from the same supertype it does not guarantee that those HLA will present the peptide in the same conformation, or that T cells will be able to recognise the peptide in the context of multiple HLA. For example, closely related HLA-B*42:01 and HLA-B*81:01 (belonging to the HLA-B7 supertype) both present the HIV derived Gag_{180–188} epitope (TPQDLNTML, TL9) but in very distinct conformations due to polymorphisms in the cleft [38] (Figure 2B). This is associated with different outcome between the HLA-B*42:01+ and HLA-B*81:01+ HIV+ patients [39]. Another example is HLA-B*37:01, HLA-B*18:01 and HLA-B*44:05 (belonging to the HLA-B44 supertype) each able to present the influenza epitope NP_{338–346} but in distinct conformations due to polymorphisms within the binding cleft [40] (Figure 2C). Here again, T cell activation was very different as the NP_{338–346} peptide is immunogenic only in the context of HLA-B*37:01 molecule. Therefore, there is no direct link between HLA molecules from the same supertype, binding the same peptide, and their ability to activate an immune response.

MHC & peptide match prediction tools

To overcome a growing need for HLA peptide restriction, prediction software were made available online, which could predict binding between peptides and HLA molecules [41–46]. These early prediction tools devised algorithms that used peptide and HLA sequences to predict binding affinity based on peptide affinity data. Today, prediction tools (for e.g. NetMHCpan 4.1) combine both binding affinity data with mass spectrometry eluted peptides using machine learning strategies to create more powerful and accurate tools for prediction [5,47,48]. Earlier versions of MHC prediction tools struggled with availability of data [41,49], yet current versions benefit from large amounts of mass spectrometry data extracted from multiple HLA molecules simultaneously, where specificity for single HLAs is later determined [50,51].

MHC prediction embodies a bottom-up approach (peptide-based data), where numerous datasets of peptide binders from each HLA are used to predict binding. This contrasts with earlier concepts of superotypes, that embody a top-down approach (HLA sequence-based), where HLA amino acid sequence and polymorphisms determine peptide binders and thus supertype classification. Today, MHC prediction tools have found success in physicochemical approaches in T cell-based therapies such as peptide vaccine design that target specific HLAs [52,53], even though this was the proposed intention of HLA superotypes classification [10].

Another approach is to use epitope databases such as Immune Epitope Database (IEDB, www.iedb.org) to compile a list of peptides known to bind a given HLA, which also has the advantage of having all the information regarding immunogenicity of the peptides (T cell activation). This list of peptides can then be curated and a peptide binding motif can be generated from known and experimentally verified epitopes, and not just peptide binders. Of course, the limitation is that only a few HLA molecules are well represented in the database, and that the number of epitopes (peptides able to activate T cells) is less than the number of peptides able to bind a given HLA molecule [54].

Comparison between structure-based and sequence based (binding data) methods

One of the initial hurdles for peptide prediction tools was the limitation of binding data [41,49]. However, as accessibility to both these became more widespread and their data more accessible, these results were able to be combined into improving HLA prediction software. Likewise, the number of pHLA structures has increased within the past two decades [8]. These pHLA structures provide empirical and definitive data of peptide restriction by HLAs displaying, in plain view, interactions within each pocket that define specificity [55].

As structural data continues to increase, so too can these data be used to train prediction tools to become more accurate [56,57]. Although knowledge of the specific interactions that influence peptide binding is expected to improve binding predictions [58,59], the development of structure-based methods have been slow relative to sequence based or binding data methods (reviewed in [56]).

Here, we investigated 14 HLA allotypes most frequently studied in structural biology or are a reference HLA within their supertype and compare their anchored residues with pocket specificities detailed in motif viewer [60] (Table 3). For HLA-A*02:01, 204 crystal structures have been solved in complex with various peptides free

Table 3 Peptide-HLA structures with unfavoured primary anchor residues

Allele	Peptide	Unfavoured Primary Anchor		PDB ID
		Pocket B	Pocket F	
A*0101	AQDIYRASYY	Q		6AT9
A*0201	GVWIRTPPA	V		5E00
A*0201	LFGYPVYV	F		1DUY
A*0201	GVYDGREHTV	V		1I4F
A*0201	FAPGFFPYL	A		1I7R
A*0201	SLLMWITQC		C	1S9W
A*0201	SLLMWITQS		S	1S9Y
A*0201	ITDQVPFSV	T		1TVB
A*0201	EAAGIGILTV	A		2GT9
A*0201	LAGIGILTV	A		2GTW
A*0201	AAGIGILTV	A		2GUO
A*0201	RQASLSISV	Q		3BGM
A*0201	RQASIELPSM	Q	M	3BH8
A*0201	RTYSGPMNKV	T		3BH9
A*0201	RVASPTSGV	V		3FQW
A*0201	RVASPTSGV	V		3FQX
A*0201	LGYGFVNYI	G		3PWL
A*0201	AAGIGILTV	A		3QFD
A*0201	LACFVLA AV	A		3TO2
A*0201	KVAELVHFL	V		3V5D
A*0201	KVAEIVHFL	V		3V5H
A*0201	KVAELVWFL	V		3V5K
A*0201	FATGIGITV	A		4GKN
A*0201	VCWGELMNL	C		4K7F
A*0201	RQASLSISV	Q		4NNX
A*0201	RQASLSISV	Q		4NNY
A*0201	RQASIELPSMAV	Q		4NO0
A*0201	RQASIELPSMAV	Q		4NO2
A*0201	RQISQDVKL	Q		4NO3
A*0201	RQISQDVKL	Q		4NO5
A*0201	GSDARFLRGYHLYA	S		4OV5
A*0201	FLNKDLEVDGHFVTM		M	4U6Y
A*0201	AQWGPDPAAA	Q		5C0D
A*0201	YQFGPDFPIA	Q		5C0E
A*0201	RQWGPDPAAV	Q		5C0F
A*0201	RQFGPDFPTI	Q		5C0I
A*0201	RQFGPDWIVA	Q		5C0J
A*0201	FVLELEPEWTV	V		5D9S
A*0201	FVLELEPEWTVK	V		5DDH
A*0201	MVWGPDPPLYV	V		5N1Y
A*0201	CVNGSCFTV	V		5SWQ
A*0201	GVWIRTPTA	V		5WSH
A*0201	MMWDRGLGMM		M	6AMT
A*0201	RTFSPTYGL	T		6EWO
A*0201	AVYDGREHTV	V		6TRN
A*0201	GVYDGREHTV	V		6TRO
A*0201	LQLPQGTTL	Q		7KGR
B*0801	HSKKKCEDEL	S		4QRQ
B*5701	RVKEYQHLW	V		6BXP
B*5701	RVAQLEQVYI	V	I	3VRI
B*5701	LTTKLTNTNI		I	3VRJ

Peptides shown in red have both pockets B and F with unfavoured primary anchor residues.

of T cell receptors (192 with complete peptide sequences). According to motif viewer [60], HLA-A*02:01 prefers L, M, or I in pocket B and V, L, I, or A in pocket F (Figure 2D). The majority of peptides conformed to these HLA-A*02:01 pocket B and F anchor preferences. For HLA-A*02:01, we found 46 out of 204 structures (23%) had bound peptides with an unpreferred anchor residue in pocket B or F. The majority of these (42/46) tolerated an unfavoured anchor at pocket B. Additionally, only one structure with both B and F pockets with unfavoured anchor residues was found for HLA-A*02:01. This structure showed HLA-A*02:01 in complex with a 10mer phosphopeptide (RQApSIELPSM) (PDB ID: 3BH8[61]), where P4-S is phosphorylated (Figure 2E). The crystal structures of several other phosphorylated P4-S sequences similar to this structure show a preference for a P1-R/P4-S motif as both P1-R and P4-S-phosphorylated side chains interact with each other and also form a network of interactions with W167 and R65 [61]. P2-Q fit poorly into the highly hydrophobic B pocket and was shown to form water-mediated contacts with Y99, whilst pushing the H70 side chain away from pocket B (both are key residues for pocket B specificity). Whilst P Ω -M caused the peptide main chain to be elevated to fit methionine's longer side chain into the smaller frame of HLA-A*02:01's F pocket. The authors hypothesised that even though P2-Q and P Ω -M are sub-optimally fit into pockets B and F, respectively, and would normally incur energetic penalties relative to more optimal anchor residues, these penalties in binding and peptide presentation are offset by phosphate-mediated interactions stabilising HLA-A*02:01. Interestingly, the peptide affinity between HLA-A*02:01 and the phosphorylated peptide is 159-fold stronger than that of the non-phosphorylated peptide sequence (i.e. RQASIELPSM) [61]. Mutations to P1 peptide or HLA R65 were shown to substantially decrease peptide binding affinity, demonstrating that these interactions were indeed contributing to the complex's high affinity [61].

Another outlier with unfavoured primary anchors was a self-peptide derived from ribonucleoprotein Sm D3, RVAQLEQVYI, in complex with HLA-B*57:01 (Table 3). Interestingly, this crystal structure was solved with abacavir, an antiretroviral used to treat HIV, binding into pockets C, D, E and F of HLA-B*57:01 (PDB ID: 3VRI [62]). Although abacavir made several contacts with HLA-B*57:01 pockets, there were limited interactions with abacavir and the peptide (only at P Ω -I), with P6-E and P7-Q peptide residues adopting a raised conformation outside of the binding cleft [62] (Figure 2F). Abacavir hypersensitivity syndrome is a HLA-associated drug reaction that exclusively affects individuals expressing the HLA-B*57:01 allele [63,64]. The abacavir-bound HLA-B*57:01 thus presents a new repertoire of peptides to the immune system, causing self-reactivity only in HLA-B*57:01+ individuals but not in HLA-B*57:02/03/11+ or HLA-B*58:01+ [62,65]. This highlights the allo-type specificity of this drug hypersensitivity, where even single amino acid polymorphisms can result in altered HLA-drug interactions. Although this has been seen prominently in abacavir hypersensitivity syndrome, other drugs can cause similar reactions [66]. For example, another HIV antiretroviral drug, Nevirapine causes hypersensitivity reactions associated in a HLA-C cluster sharing a similar F pocket to HLA-C*04:01, which could indicate potential drug binding [67].

Therefore, there is a lot of plasticity in the way HLA pockets can bind residues or other small molecules, altogether providing insight into the broad repertoire of ligands that T cells recognise.

Conclusion

HLAs are a key player in the immune system, being the primary target of T cells. Gaining a better understanding of their peptide binding capabilities will help inform researchers on the quality of the immune response to any given pathogen. The initial description of the HLA supertype families helped group a large and growing number of HLA molecules. However, as more data became available and the number of newly discovered alleles grew substantially, this classification needed to evolve and be revisited to remain relevant and useful. Today, over 22 000 HLA-I alleles have been discovered and classifying all these alleles into superotypes would be an enormous feat.

While the current tools and data helped build up our current knowledge and give us a clear picture of how HLA molecules bind peptides, we are still far from being able to properly predict which peptides can successfully be presented by any given HLA. A step further will be to correlate peptide binding and immunogenicity, to not only enable our understanding of which peptide can be presented, but also which peptide bound HLAs will activate T cells.

This predictive ability will have enormous potential to help design better drugs to avoid drug hypersensitivity [62–66], as well as rapidly and accurately predict epitopes for newly emerging pathogens. A clear application that could help the current research of viral immunity, would be the prediction of SARS-CoV-2 peptides able to bind HLA molecules. A lot of the initial work on SARS-CoV-2 used peptide binding prediction, or prior

knowledge of closely related viruses (i.e. SARS), to select relevant peptides to be studied. However, some early predictions failed to correctly identify HLA restriction for some peptides [68] or are only accurate at predicting a handful of the most common and well-studied HLA allotypes. Therefore, prediction for a larger number of HLA molecules would be highly informative and would help focus the study on selected peptides.

The holy grail of peptide prediction would be the ability to predict dominant immunogenic epitopes for a given HLA. Hopefully this can be achieved with widespread implementation and improvement of computational approaches using all data available (prediction, peptide affinity, mass spectrometry, T cell activation and pHLA structures). This information will be highly relevant for therapeutics such as peptide-based vaccine, and even developing personalized T cell-based therapy against pathogenic infections or cancer.

Perspectives

- Understanding peptide binding specificity to HLA is key for developing T cell-based therapies such as vaccines.
- HLA grouping into supertypes is a quick and easy way to draw associations between HLA and peptide, but falls short to correctly help predicting peptide binding.
- Structural data integrated into the current predictive algorithm will improve peptide prediction and HLA association, with future algorithms able to predict immunogenicity as well.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

This work was supported by the Australian Research Council (ARC), and National Health and Medical Research Council (NHMRC). A.T.N. is supported by a Monash Biomedicine Discovery Institute PhD Scholarship from Monash University, and an Australian Institute of Nuclear Science and Engineering (AINSE) Ltd. Postgraduate Research Award (PGRA) from the Australian Nuclear Science and Technology organisation (ANSTO), C.S. is supported by an AINSE ECR grant, and S.G. is supported by an NHMRC Senior Research Fellowship (#1159272).

Open Access

Open access for this article was enabled by the participation of La Trobe University in an all-inclusive *Read & Publish* pilot with Portland Press and the Biochemical Society under a transformative agreement with CAUL.

Author Contributions

A.T.N., C.S., and S.G. have curated, analyses the data, written and approved the manuscript.

Acknowledgements

The authors thanks all their colleagues working in the field of peptide-MHC.

Abbreviations

HLA, Human leukocyte antigens; IEDB, Immune Epitope Database; MHC, major histocompatibility complex; PDB, Protein Data Bank.

References

- 1 Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. and Wiley, D.C. (1987) Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506–512 <https://doi.org/10.1038/329506a0>
- 2 Josephs, T.M., Grant, E.J. and Gras, S. (2017) Molecular challenges imposed by MHC-I restricted long epitopes on T cell immunity. *Biol. Chem.* **398**, 1027–1036 <https://doi.org/10.1515/hsz-2016-0305>

- 3 Jardtzyk, T.S., Brown, J.H., Gorga, J.C., Stern, L.J., Urban, R.G., Chi, Y.I. et al. (1994) Three-dimensional structure of a human class II histocompatibility molecule complexed with superantigen. *Nature* **368**, 711–718 <https://doi.org/10.1038/368711a0>
- 4 Sercinoglu, O. and Ozbek, P. (2020) Sequence-structure-function relationships in class I MHC: a local frustration perspective. *PLoS ONE* **15**, e0232849 <https://doi.org/10.1371/journal.pone.0232849>
- 5 Gfeller, D. and Bassani-Sternberg, M. (2018) Predicting antigen presentation-what could we learn from a million peptides? *Front. Immunol.* **9**, 1716 <https://doi.org/10.3389/fimmu.2018.01716>
- 6 van Deutekom, H.W. and Kesmir, C. (2015) Zooming into the binding groove of HLA molecules: which positions and which substitutions change peptide binding most? *Immunogenetics* **67**, 425–436 <https://doi.org/10.1007/s00251-015-0849-y>
- 7 Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H. et al. (2000) The protein data bank. *Nucleic Acids Res.* **28**, 235–242 <https://doi.org/10.1093/nar/28.1.235>
- 8 Szeto, C., Lobos, C.A., Nguyen, A.T. and Gras, S. (2020) TCR recognition of peptide-MHC-I: rule makers and breakers. *Int. J. Mol. Sci.* **22**, 68 <https://doi.org/10.3390/ijms22010068>
- 9 Saper, M.A., Bjorkman, P.J. and Wiley, D.C. (1991) Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol.* **219**, 277–319 [https://doi.org/10.1016/0022-2836\(91\)90567-P](https://doi.org/10.1016/0022-2836(91)90567-P)
- 10 Sette, A. and Sidney, J. (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* **50**, 201–212 <https://doi.org/10.1007/s002510050594>
- 11 Chujoh, Y., Sobao, Y., Miwa, K., Kaneko, Y. and Takiguchi, M. (1998) The role of anchor residues in the binding of peptides to HLA-A*1101 molecules. *Tissue Antigens* **52**, 501–509 <https://doi.org/10.1111/j.1399-0039.1998.tb03080.x>
- 12 Sidney, J., Peters, B., Frahm, N., Brander, C. and Sette, A. (2008) HLA class I supertypes: a revised and updated classification. *BMC Immunol.* **9**, 1 <https://doi.org/10.1186/1471-2172-9-1>
- 13 Rowntree, L.C., Nguyen, T.H.O., Halim, H., Purcell, A.W., Rossjohn, J., Gras, S. et al. (2018) Inability to detect cross-reactive memory T cells challenges the frequency of heterologous immunity among common viruses. *J. Immunol.* **200**, 3993–4003 <https://doi.org/10.4049/jimmunol.1800010>
- 14 Remesh, S.G., Andreatta, M., Ying, G., Kaefer, T., Nielsen, M., McMurtrey, C. et al. (2017) Unconventional peptide presentation by major histocompatibility complex (MHC) class I allele HLA-A*02:01: breaking confinement. *J. Biol. Chem.* **292**, 5262–5270 <https://doi.org/10.1074/jbc.M117.776542>
- 15 Motozono, C., Pearson, J.A., De Leenheer, E., Rizkallah, P.J., Beck, K., Trimby, A. et al. (2015) Distortion of the major histocompatibility complex class I binding groove to accommodate an insulin-derived 10-Mer peptide. *J. Biol. Chem.* **290**, 18924–18933 <https://doi.org/10.1074/jbc.M114.622522>
- 16 Le Nours, J., Shahine, A. and Gras, S. (2018) Molecular features of lipid-based antigen presentation by group 1 CD1 molecules. *Semin. Cell Dev. Biol.* **84**, 48–57 <https://doi.org/10.1016/j.semcdb.2017.11.002>
- 17 Ruppert, J., Kubo, R.T., Sidney, J., Grey, H.M. and Sette, A. (1994) Class I MHC-peptide interaction: structural and functional aspects. *Behring Inst. Mitt.* **94**, 48–60 PMID:7998914
- 18 Hensen, L., Illing, P.T., Bridie Clemens, E., Nguyen, T.H.O., Koutsakos, M., van de Sandt, C.E. et al. (2021) CD8(+) t cell landscape in indigenous and non-indigenous people restricted by influenza mortality-associated HLA-A*24:02 allomorph. *Nat. Commun.* **12**, 2931 <https://doi.org/10.1038/s41467-021-23212-x>
- 19 del Guercio, M.F., Sidney, J., Hermanson, G., Perez, C., Grey, H.M., Kubo, R.T. et al. (1995) Binding of a peptide antigen to multiple HLA alleles allows definition of an A2-like supertype. *J. Immunol.* **154**, 685–693 PMID:7529283
- 20 Sidney, J., Grey, H.M., Kubo, R.T. and Sette, A. (1996) Practical, biochemical and evolutionary implications of the discovery of HLA class I supermotifs. *Immunol. Today* **17**, 261–266 [https://doi.org/10.1016/0167-5699\(96\)80542-1](https://doi.org/10.1016/0167-5699(96)80542-1)
- 21 Sylvester-Hvid, C., Nielsen, M., Lamberth, K., Roder, G., Justesen, S., Lundegaard, C. et al. (2004) SARS CTL vaccine candidates; HLA supertype-, genome-wide scanning and biochemical validation. *Tissue Antigens* **63**, 395–400 <https://doi.org/10.1111/j.0001-2815.2004.00221.x>
- 22 Reche, P.A. and Reinherz, E.L. (2005) PEPVAC: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands. *Nucleic Acids Res.* **33**, W138–W142 <https://doi.org/10.1093/nar/gki357>
- 23 Khan, A.M., Miotto, O., Heiny, A.T., Salmon, J., Srinivasan, K.N., Nascimento, E.J. et al. (2006) A systematic bioinformatics approach for selection of epitope-based vaccine targets. *Cell Immunol.* **244**, 141–147 <https://doi.org/10.1016/j.cellimm.2007.02.005>
- 24 Frahm, N., Yusim, K., Suscovich, T.J., Adams, S., Sidney, J., Hraber, P. et al. (2007) Extensive HLA class I allele promiscuity among viral CTL epitopes. *Eur. J. Immunol.* **37**, 2419–2433 <https://doi.org/10.1002/eji.200737365>
- 25 Zhang, G.L., Khan, A.M., Srinivasan, K.N., August, J.T. and Brusica, V. (2005) MULTIPRED: a computational system for prediction of promiscuous HLA binding peptides. *Nucleic Acids Res.* **33**, W172–W179 <https://doi.org/10.1093/nar/gki452>
- 26 Axelsson-Robertson, R., Weichold, F., Sizemore, D., Wulf, M., Skeiky, Y.A., Sadoff, J. et al. (2010) Extensive major histocompatibility complex class I binding promiscuity for mycobacterium tuberculosis TB10.4 peptides and immune dominance of human leucocyte antigen (HLA)-B*0702 and HLA-B*0801 alleles in TB10.4 CD8T-cell responses. *Immunology* **129**, 496–505 <https://doi.org/10.1111/j.1365-2567.2009.03201.x>
- 27 Rao, X., Hoof, I., Costa, A.I., van Baarle, D. and Kesmir, C. (2011) HLA class I allele promiscuity revisited. *Immunogenetics* **63**, 691–701 <https://doi.org/10.1007/s00251-011-0552-6>
- 28 Harjanto, S., Ng, L.F. and Tong, J.C. (2014) Clustering HLA class I superfamilies using structural interaction patterns. *PLoS ONE* **9**, e86655 <https://doi.org/10.1371/journal.pone.0086655>
- 29 Lund, O., Nielsen, M., Kesmir, C., Petersen, A.G., Lundegaard, C., Worning, P. et al. (2004) Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics* **55**, 797–810 <https://doi.org/10.1007/s00251-004-0647-4>
- 30 Hertz, T. and Yanover, C. (2007) Identifying HLA supertypes by learning distance functions. *Bioinformatics* **23**, e148–e155 <https://doi.org/10.1093/Bioinformatics/btl324>
- 31 Doytchinova, I.A., Guan, P. and Flower, D.R. (2004) Identifying human MHC supertypes using bioinformatic methods. *J. Immunol.* **172**, 4314–4323 <https://doi.org/10.4049/jimmunol.172.7.4314>
- 32 Mukherjee, S., Warwicker, J. and Chandra, N. (2015) Deciphering complex patterns of class-I HLA-peptide cross-reactivity via hierarchical grouping. *Immunol. Cell Biol.* **93**, 522–532 <https://doi.org/10.1038/icb.2015.3>

- 33 Mumtaz, S., Nabney, I.T. and Flower, D.R. (2017) Scrutinizing human MHC polymorphism: Supertype analysis using Poisson-Boltzmann electrostatics and clustering. *J. Mol. Graph. Model.* **77**, 130–136 <https://doi.org/10.1016/j.jmkgm.2017.07.033>
- 34 Trachtenberg, E., Korber, B., Sollars, C., Kepler, T.B., Hraber, P.T., Hayes, E. et al. (2003) Advantage of rare HLA supertype in HIV disease progression. *Nat. Med.* **9**, 928–935 <https://doi.org/10.1038/nm893>
- 35 Eickhoff, C.S., Terry, F.E., Peng, L., Meza, K.A., Sakala, I.G., Van Aartsen, D. et al. (2019) Highly conserved influenza T cell epitopes induce broadly protective immunity. *Vaccine* **37**, 5371–5381 <https://doi.org/10.1016/j.vaccine.2019.07.033>
- 36 Lazaryan, A., Song, W., Lobashevsky, E., Tang, J., Shrestha, S., Zhang, K. et al. (2010) Human leukocyte antigen class I supertypes and HIV-1 control in African Americans. *J. Virol.* **84**, 2610–2617 <https://doi.org/10.1128/JVI.01962-09>
- 37 Goulder, P.J. and Walker, B.D. (2012) HIV and HLA class I: an evolving relationship. *Immunity* **37**, 426–440 <https://doi.org/10.1016/j.immuni.2012.09.005>
- 38 Klooverpris, H.N., Cole, D.K., Fuller, A., Carlson, J., Beck, K., Schauenburg, A.J. et al. (2015) A molecular switch in immunodominant HIV-1-specific CD8T-cell epitopes shapes differential HLA-restricted escape. *Retrovirology* **12**, 20 <https://doi.org/10.1186/s12977-015-0149-5>
- 39 Geldmacher, C., Metzler, I.S., Tovanabutra, S., Asher, T.E., Gostick, E., Ambrozak, D.R. et al. (2009) Minor viral and host genetic polymorphisms can dramatically impact the biologic outcome of an epitope-specific CD8T-cell response. *Blood* **114**, 1553–1562 <https://doi.org/10.1182/blood-2009-02-206193>
- 40 Grant, E.J., Josephs, T.M., Loh, L., Clemens, E.B., Sant, S., Bharadwaj, M. et al. (2018) Broad CD8(+) T cell cross-recognition of distinct influenza A strains in humans. *Nat. Commun.* **9**, 5427 <https://doi.org/10.1038/s41467-018-07815-5>
- 41 Nielsen, M., Lundegaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S. et al. (2007) NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE* **2**, e796 <https://doi.org/10.1371/journal.pone.0000796>
- 42 Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A. and Stevanovic, S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50**, 213–219 <https://doi.org/10.1007/s002510050595>
- 43 Reche, P.A., Glutting, J.P. and Reinherz, E.L. (2002) Prediction of MHC class I binding peptides using profile motifs. *Hum. Immunol.* **63**, 701–709 [https://doi.org/10.1016/S0198-8859\(02\)00432-9](https://doi.org/10.1016/S0198-8859(02)00432-9)
- 44 Bhasin, M., Singh, H. and Raghava, G.P. (2003) MHCBN: a comprehensive database of MHC binding and non-binding peptides. *Bioinformatics* **19**, 665–666 <https://doi.org/10.1093/bioinformatics/btg055>
- 45 Govindarajan, K.R., Kanguane, P., Tan, T.W. and Ranganathan, S. (2003) MPID: MHC-peptide interaction database for sequence-structure-function information on peptides binding to MHC molecules. *Bioinformatics* **19**, 309–310 <https://doi.org/10.1093/bioinformatics/19.2.309>
- 46 Reche, P.A., Zhang, H., Glutting, J.P. and Reinherz, E.L. (2005) EPIMHC: a curated database of MHC-binding peptides for customized computational vaccinology. *Bioinformatics* **21**, 2140–2141 <https://doi.org/10.1093/bioinformatics/bti269>
- 47 Zhao, T., Cheng, L., Zang, T. and Hu, Y. (2019) Peptide-major histocompatibility complex class I binding prediction based on deep learning with novel feature. *Front. Genet.* **10**, 1191 <https://doi.org/10.3389/fgene.2019.01191>
- 48 Luo, H., Ye, H., Ng, H.W., Shi, L., Tong, W., Mendrick, D.L. et al. (2015) Machine learning methods for predicting HLA-peptide binding activity. *Bioinform. Biol. Insights* **9**, 21–29 <https://doi.org/10.4137/BBI.S29466>
- 49 Nielsen, M., Lundegaard, C., Blicher, T., Peters, B., Sette, A., Justesen, S. et al. (2008) Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS Comput. Biol.* **4**, e1000107 <https://doi.org/10.1371/journal.pcbi.1000107>
- 50 Abelin, J.G., Keskin, D.B., Sarkizova, S., Hartigan, C.R., Zhang, W., Sidney, J. et al. (2017) Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity* **46**, 315–326 <https://doi.org/10.1016/j.immuni.2017.02.007>
- 51 Reynisson, B., Alvarez, B., Paul, S., Peters, B. and Nielsen, M. (2020) NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W554 <https://doi.org/10.1093/nar/gkaa379>
- 52 Schubert, B., Lund, O. and Nielsen, M. (2013) Evaluation of peptide selection approaches for epitope-based vaccine design. *Tissue Antigens* **82**, 243–251 <https://doi.org/10.1111/tan.12199>
- 53 Wang, S., Guo, L., Liu, D., Liu, W. and Wu, Y. (2016) HLAstupe: an integrated database of HLA supertype-specific epitopes to aid in the development of vaccines with broad coverage of the human population. *BMC Immunol.* **17**, 17 <https://doi.org/10.1186/s12865-016-0156-x>
- 54 Koutsakos, M., Illing, P.T., Nguyen, T.H.O., Mifsud, N.A., Crawford, J.C., Rizzetto, S. et al. (2019) Human CD8(+) T cell cross-reactivity across influenza A, B and C viruses. *Nat. Immunol.* **20**, 613–625 <https://doi.org/10.1038/s41590-019-0320-6>
- 55 Tong, J.C., Tan, T.W. and Ranganathan, S. (2007) In silico grouping of peptide/HLA class I complexes using structural interaction characteristics. *Bioinformatics* **23**, 177–183 <https://doi.org/10.1093/bioinformatics/btl563>
- 56 Liao, W.W. and Arthur, J.W. (2011) Predicting peptide binding affinities to MHC molecules using a modified semi-empirical scoring function. *PLoS ONE* **6**, e25055 <https://doi.org/10.1371/journal.pone.0025055>
- 57 Antunes, D.A., Devaurs, D., Moll, M., Lizée, G. and Kaviraki, L.E. (2018) General prediction of peptide-MHC binding modes using incremental docking: a proof of concept. *Sci. Rep.* **8**, 4327 <https://doi.org/10.1038/s41598-018-22173-4>
- 58 Zhang, H., Wang, P., Papangelopoulos, N., Xu, Y., Sette, A., Bourne, P.E. et al. (2010) Limitations of Ab initio predictions of peptide binding to MHC class II molecules. *PLoS ONE* **5**, e9272 <https://doi.org/10.1371/journal.pone.0009272>
- 59 Jojic, N., Reyes-Gomez, M., Heckerman, D., Kadie, C. and Schueler-Furman, O. (2006) Learning MHC I-peptide binding. *Bioinformatics* **22**, e227–e235 <https://doi.org/10.1093/bioinformatics/btl255>
- 60 Rapin, N., Hoof, I., Lund, O. and Nielsen, M. (2008) MHC motif viewer. *Immunogenetics* **60**, 759–765 <https://doi.org/10.1007/s00251-008-0330-2>
- 61 Mohammed, F., Cobbold, M., Zarling, A.L., Salim, M., Barrett-Wilt, G.A., Shabanowitz, J. et al. (2008) Phosphorylation-dependent interaction between antigenic peptides and MHC class I: a molecular basis for the presentation of transformed self. *Nat. Immunol.* **9**, 1236–1243 <https://doi.org/10.1038/ni.1660>
- 62 Illing, P.T., Vivian, J.P., Dudek, N.L., Kostenko, L., Chen, Z., Bharadwaj, M. et al. (2012) Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. *Nature* **486**, 554–558 <https://doi.org/10.1038/nature11147>
- 63 Mallal, S., Nolan, D., Witt, C., Masel, G., Martin, A.M., Moore, C. et al. (2002) Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor Abacavir. *Lancet* **359**, 727–732 [https://doi.org/10.1016/S0140-6736\(02\)07873-X](https://doi.org/10.1016/S0140-6736(02)07873-X)

- 64 Hetherington, S., Hughes, A.R., Mosteller, M., Shortino, D., Baker, K.L., Spreen, W. et al. (2002) Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. *Lancet* **359**, 1121–1122 [https://doi.org/10.1016/S0140-6736\(02\)08158-8](https://doi.org/10.1016/S0140-6736(02)08158-8)
- 65 Chessman, D., Kostenko, L., Lethborg, T., Purcell, A.W., Williamson, N.A., Chen, Z. et al. (2008) Human leukocyte antigen class I-restricted activation of CD8+ T cells provides the immunogenetic basis of a systemic drug hypersensitivity. *Immunity* **28**, 822–832 <https://doi.org/10.1016/j.immuni.2008.04.020>
- 66 Illing, P.T., Mifsud, N.A. and Purcell, A.W. (2016) Allotype specific interactions of drugs and HLA molecules in hypersensitivity reactions. *Curr. Opin. Immunol.* **42**, 31–40 <https://doi.org/10.1016/j.coi.2016.05.003>
- 67 Pavlos, R., McKinnon, E.J., Ostrov, D.A., Peters, B., Buus, S., Koelle, D. et al. (2017) Shared peptide binding of HLA class I and II alleles associate with cutaneous nevirapine hypersensitivity and identify novel risk alleles. *Sci. Rep.* **7**, 8653 <https://doi.org/10.1038/s41598-017-08876-0>
- 68 Szeto, C., Chatzileontiadou, D.S.M., Nguyen, A.T., Sloane, H., Lobos, C.A., Jayasinghe, D. et al. (2021) The presentation of SARS-CoV-2 peptides by the common HLA-A(*):02:01 molecule. *iScience* **24**, 102096 <https://doi.org/10.1016/j.isci.2021.102096>
- 69 Thomsen, M.C. and Nielsen, M. (2012) Seq2logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res.* **40**, W281–W287 <https://doi.org/10.1093/nar/gks469>