

## Dataset Paper

# A Dataset of Experimental HLA-B\* 2705 Peptide Binding Affinities

Valerie A. Walshe,<sup>1</sup> Channa K. Hattotuwigama,<sup>2</sup>  
Irina A. Doytchinova,<sup>3</sup> and Darren R. Flower<sup>4</sup>

<sup>1</sup> Centre for Emergency Preparedness and Response, Porton Down, Salisbury, Wiltshire SP4 0JG, UK

<sup>2</sup> GlaxoSmithKline, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, UK

<sup>3</sup> Faculty of Pharmacy, Medical University of Sofia, 2 Dunav Street, 1000 Sofia, Bulgaria

<sup>4</sup> Aston Pharmacy School, School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK

Correspondence should be addressed to Darren R. Flower; [d.r.flower@aston.ac.uk](mailto:d.r.flower@aston.ac.uk)

Received 6 November 2013; Accepted 10 February 2014; Published 27 April 2014

Academic Editors: T. Šarić and C.-W. Tung

Copyright © 2014 Valerie A. Walshe et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

T-cell epitopes form the basis of many vaccines, diagnostics, and reagents. Current methods for the *in silico* identification of T-cell epitopes rely, in the main, on the accurate quantitative prediction of peptide-Major Histocompatibility Complex (pMHC) affinity using data-driven computational approaches. Here, we describe a dataset of experimentally determined pMHC binding affinities for the problematic human class I allele HLA-B\* 2705. Using an in-house, FACS-based, MHC stabilization assay, we measured binding of 223 peptides. This dataset includes both nonbinding and binding peptides, with measured affinities (expressed as  $-\log_{10}$  of the half-maximal binding level) ranging from 1.2 to 7.4. This dataset should provide a useful independent benchmark for new and existing methods for predicting peptide binding to HLA-B\* 2705.

## 1. Introduction

Products of the Major Histocompatibility Complex (MHC) play a fundamental role in regulating immune responses. T cells recognise antigen as peptide fragments bound by MHC molecules, a process requiring initial antigen degradation through complex proteolytic digestion prior to formation of a binary complex. The biological role of MHC proteins is thus to bind peptides and “present” these at the cell surface for inspection by T-cell antigen receptors (TCRs) [1]. Class I molecules are composed of a heavy chain in complex with  $\beta_2$ -microglobulin. The MHC-peptide-binding site consists of a  $\beta$ -sheet, forming the base, flanked by two  $\alpha$ -helices, which together form a narrow cleft or groove accommodating bound peptides. The principal difference between class I and class II MHCs is the structure of the peptide binding groove: this is constrained to bind 8–11 amino acid peptides in class

I, although it has become clear recently that much longer peptides can also bind to class I MHCs [2].

Predictive models of peptide-MHC binding affinity have become important components of modern computational immunovaccinology [3]. Previously, such approaches have been built around relatively uncomplicated classification methods but have now largely given way to quantitative regression-based approaches [4]. Immunoinformatics, a newly emergent subdiscipline of bioinformatics, which addresses informatic problems within immunology, uses QSAR technology to tackle the crucial issue of epitope prediction [5]. As high-throughput biology reveals the genomic sequences of pathogenic bacteria, viruses, and parasites, such prediction will become increasingly important in the postgenomic discovery of novel vaccines, reagents, and diagnostics. In order to better understand the sequence dependence of peptide-MHC binding of the mouse MHC, we have

previously explored the amino acid preferences of various human and mouse alleles. Computer models that simulate peptide binding to MHC are useful for selecting candidate T-cell epitopes since they minimise the number of experiments required for their identification [6].

The human MHC allele HLA-B\*2705 is one of the most intriguing and perplexing molecules of its kind. Possession of HLA-B\*2705 confers susceptibility to the inflammatory arthritic disorder ankylosing spondylitis (AS) and other related arthropathies, of which arthritis is the best known disorder. Despite thirty years of effort, the specific mechanism which underlies this association remains unclear. Other properties unrelated to its role in antigen presentation may likewise be important in disease pathogenesis. The role of HLA-B27-peptide binding remains critical to any convincing and compelling exegesis of spondyloarthropathies, despite the many alternative explanations provided for HLA-B27 subtypes' involvement in the pathogenesis of spondyloarthropathies [7]: (1) oxidative misfolding of HLA-B27, (2) cell surface expression of HLA-B27 homodimers, (3)  $\beta_2$ -microglobulin-free and peptide-free HLA-B27 heavy chain expression, (4) HLA-B27 modulated ERAAP and tapasin function, and (5)  $\beta_2$ -microglobulin overexpression and/or deposition, amongst many others. Each of these mechanisms, or any combination thereof, may contribute causally to the emergence of spondyloarthropathy, yet each is itself related to peptide binding. For example, peptide specificity determines the stability of the trimeric pMHC complex and also T-cell recognition, which underlies the arthritogenic peptide hypothesis. HLA-B27's ability to bind a stable peptide repertoire in the endoplasmic reticulum influences the folding rate influencing, with  $\beta_2$ -microglobulin supply, the extent of misfolded heavy chain accumulating in the ER. This then underlies the misfolding hypothesis and so on for other explanatory mechanisms [8].

The large-scale identification of HLA-B\*2705 ligands and, to a lesser extent, other HLA-B27 subtypes [9, 10] has been motivated by the supposed role of peptides in the strong association of HLA-B27 with spondyloarthropathies [11]. HLA-B27 subtypes account for about one-third of genetic susceptibility to spondyloarthropathies. Up to 70 HLA-B27 subtypes have been reported globally, with a decreasing north-south gradient in frequency, which is, interestingly, the essential opposite of endemic malaria. From a variety of studies, which include X-ray structure determination and peptide sequencing, the peptide specificity of HLA-B27 has been adumbrated in some detail. The principal features of HLA-B27 ligands are well known. Peptides that bind to HLA-B\*2705 are typically 8–11 amino acids in length and have two main anchor residues at positions 2 and the C-terminal end [12]. There seems to be a near absolute requirement for arginine at peptide position 2, the *de facto* main anchor residue. The peptide C-terminus is thought to be the second most important anchor. Residues at this position, whose binding is modulated by subtype polymorphism, are seemingly restricted to basic, aliphatic, and aromatic amino acids. Other secondary anchors (P1, P3, and P7) have less significant restrictions in residue usage. Other positions (P4, P5, P6, and P8) are considered as nonanchors. These positions have a very relaxed selectivity

and no real residue bias that has been reported. Generally speaking, the presence of anchors is deemed to be necessary, but not sufficient, for high affinity binding. Prominent roles for several other positions 1, 3, and 7, so-called secondary anchor residues are also widely reported [13].

Many studies have, over time, attempted to develop reliable binding assays for HLA-B27. Compared to assays for the most well studied of all human MHC alleles, HLA-A\*0201, binding assays for HLA-B27 have proven to be both unreliable and, from a bioinformatics perspective, also nonpredictable. It is not clear, however, whether this arises from purely technical problems with the assays used or from more fundamental problems of interpreting the underlying molecular mechanism of peptide binding to HLA-B27. In this study, we determined the affinities of 223 peptides binding to the MHC class I allele HLA-B\*2705 using an in-house, fluorescence-activated cell sorting- (FACS-) based, MHC stabilization assay [14–17].

## 2. Methodology

Peptides were selected for study from two sources: extant legacy peptides and designed peptides (data not shown). Peptides used were obtained from collaborators (legacy peptides) or ordered either from Mimotopes (Pensby, UK) or from the Institute of Animal Health (Compton) in-house peptide synthesis service (designed peptides). Peptide binding to HLA-B\*2705 was assessed using a FACS-based MHC stabilization assay [14], with modifications as described elsewhere [17]. Briefly, T2 cells were incubated in 96-well flat-bottom plates at  $2 \times 10^5$  cells per well in a 200  $\mu$ L volume of AIM V medium (Life Technologies, Paisley, UK) with human  $\beta_2$ -microglobulin at a final concentration of 100 nM (Scipac, Sittingbourne, UK) with and without peptides at concentrations between 200 and 0.04  $\mu$ M for 16 h at 37°C. Cells were then washed and surface levels of HLA-B\*2705 were assessed by staining with a mouse anti-human HLA-B27 fluorescein isothiocyanate (FITC) Ab (Serotec). Cells were fixed at 4°C in 4% paraformaldehyde and analysed on a FACS Calibur (BD Biosciences) using CellQuest software. Results are expressed as fluorescence index (FI) values. These were calculated as the test mean fluorescence intensity (MFI) minus the no peptide isotype control MFI divided by the no peptide HLA-B\*2705-stained control MFI minus the no peptide isotype control MFI. The half-maximal binding level ( $BL_{50}$ ) which is the peptide concentration yielding the half-maximal FI of the reference peptide in each assay was calculated and presented as  $pBL_{50}$  ( $-\log BL_{50}$ ). The known HLA-B\*2705 high binder GRLTKHTKF was used as a reference peptide. We summarize, as shown in Table 1, the sequence data from column 3, in the form of an extended affinity motif, as defined by immunoinformatic analysis of this data [14], of significantly favoured and disfavoured binding residues. For certain positions (2, 3, and 5) single strongly preferred residues were seen, but for other positions (1, 4, 6, 7, 8, and 9) preferred amino acids were more diverse. Peptide positions 2 and 9 (C-terminal) are primary anchors for HLA-B\*2705 molecule [10], while positions 1, 3, 6, and 7 are considered secondary anchors [11]. Our data identified Arg at position 2 and Asn

TABLE 1: Dataset summary of peptide sequences. Anchor and nonanchor residues related to strong and weak binding.

	P1	P2	P3	P4	P5	P6	P7	P8	P9
Favoured binding	G, S	R	Q	Q, W	K	I, T	T, W	G, K	N, K
Disfavoured binding	A, K	Q, P	R	R, T	L	N, V	R, Y	R, L	R, L

A cut-off value of  $> \pm 0.1$  is applied to identify favoured and disfavoured binding amino acids.

and Lys for position 9, as anchor residues. For position 1, Gly and Ser are preferred residues; for position 3, Gln; for position 4, Gln and Trp; for position 5, Lys; for position 6, Ile and Thr; for position 7, Thr and Trp; for position 8, Gly and Lys.

### 3. Dataset Description

The dataset associated with this Dataset Paper consists of one item which is described as follows.

*Dataset Item 1 (Table).* 223 peptide-MHC binding affinity measurements determined using a FACS-based MHC stabilization assay [14] as modified previously [17]. Half-maximal binding levels ( $BL_{50}$ ) were first converted to  $\log[1/BL_{50}]$  values (or  $-\log_{10}[BL_{50}]$  or  $pBL_{50}$ ).  $pBL_{50}$  can be related to changes in the free energy of binding:  $\Delta G_{\text{bind}}\alpha - RT \ln BL_{50}$ . Three of the tested peptides were decamers, one was an octamer, and the remainder were nonameric. Of the 223 peptides, seven peptides had measured affinities greater than 7.0, with the most affine peptide, ERSGLYPQK, having an affinity of 7.4. One hundred and eight peptides had affinities between 6.0 and 7.0, forty-two peptides had affinities between 5.0 and 6.0, and twenty-nine peptides had measurable affinities below 5.0. Thirty-six peptides were nonbinders. Peptide sampling indicated that less than 25% of our peptides are available via the Immune Epitope Database (IEDB). For comparative purposes, we have also included affinities predicted using three popular and easily accessed methods: netMHC [18], SMM [19], and Comblib [20]. It may be significant that the correlation between the predicted and measured data (0.35 experimental versus netMHC, 0.06 experimental versus SMM, and 0.05 experimental versus Comblib, for 186 nonnull values) is so poor compared to the higher correlation between predicted values (0.52 netMHC versus Comblib, 0.65 netMHC versus SMM, and 0.51 SMM versus Comblib). Whether this reflects deficiencies in our data or in the predictions or in both remains an open question and a question warranting further exploration. In the table, the column Peptide Number represents the peptide identification number; 9-mer Peptide Sequence, the sequence of the tested nonameric peptides, expressed using the IUPAC 1-letter code for the 20 biogenic amino acids; 10-mer Peptide Sequence, the sequence of the tested decameric peptides, expressed using the IUPAC 1-letter code for the 20 biogenic amino acids; 8-mer Peptide Sequence, the sequence of the tested octameric peptide, expressed using the IUPAC 1-letter code for the 20 biogenic amino acids; Experimental  $pBL_{50}$  (HLA-B\*2705), the experimental affinity of peptide binding to HLA-B\*2705 expressed as half-maximal binding levels ( $BL_{50}$ ), converted to  $\log[1/BL_{50}]$  values (or  $-\log_{10}[BL_{50}]$  or  $pBL_{50}$ ). The column netMHC Predictions shows that the

affinity of peptide binding to HLA-B\*2705 was predicted using the netMHC server [18] (affinity is expressed as the predicted  $IC_{50}$  in nanomoles); IEDB:SMM Predictions shows that affinity of peptide binding to HLA-B\*2705 was predicted using the IEDB prediction server [21] implementing the SMM method [19] (affinity is expressed as the predicted  $IC_{50}$  in nanomoles); IEDB:Comblib Predictions shows that the affinity of peptide binding to HLA-B\*2705 was predicted using the IEDB prediction server [21] implementing the Comblib method [20] (virtual affinity is expressed as the percentile rank against a heterogeneous background distribution of predicted scores).

*Column 1:* Peptide Number

*Column 2:* 9-mer Peptide Sequence

*Column 3:* 10-mer Peptide Sequence

*Column 4:* 8-mer Peptide Sequence

*Column 5:* Experimental  $pBL_{50}$

*Column 6:* netMHC Predictions

*Column 7:* IEDB:SMM Predictions

*Column 8:* IEDB:Comblib Predictions

### 4. Concluding Remarks

Phenomena such as HLA-B27 misfolding and the propensity for the HLA-B27 ternary complex to dissociate have implications, both direct and indirect, for peptide binding, particularly in the cellular context. Peptide binding to HLA-B\*2705 is of undoubted importance to both autoimmune diseases and pathologies, such as spondyloarthropathies, and to infectious disease, as well as other aspects of somatic homeostasis, such as the surveillance of cancer cells. Class I MHCs normally are associated with peptide and beta(2)-microglobulin in the endoplasmic reticulum before they reach the cell surface. HLA-B27 can form homodimers through an unpaired binding site residue CYS67 and potentially also conserved structural CYS164; this may result from HLA-B27 folding kinetics, which is thought to be slower than for other class I MHCs. The possibility that there exists a complex dynamic equilibrium between a variety of HLA-B27 cysteine linked dimers and multimers, with either some nonstandard proteins reaching the cell surface or the unquantified sequestration of loaded peptide may have compromised our data, prompting the need for a fully rigorous analysis of binding peptides in the cellular context as well as the cell-free analysis of peptide binding to fully folded recombinant protein. Only through such a comparison can we hope to disentangle the intrinsic binding specificity of isolated, soluble HLA-B\*2705 from that of membrane bound

HLA-B\*2705 when influenced by other cellular components. As computer-based epitope identification becomes an ever more important component of efficient therapeutic and vaccine discovery, a fully predictive, robust, and reliable HLA-B\*2705 peptide binding model based on data such as that we report here will likewise also become a key tool in the fight against iniquitous autoimmune conditions.

## Dataset Availability

The dataset associated with this Dataset Paper is dedicated to the public domain using the CC0 waiver and is available at <http://dx.doi.org/10.1155/2014/914684/dataset>.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The authors wish to thank Professor Paul Bowness and his lab for the provision of peptides and Dr. Persephone Borrow for her kind help, advice, and assistance in undertaking this study.

## References

- [1] D. R. Flower, "Towards *in silico* prediction of immunogenic epitopes," *Trends in Immunology*, vol. 24, no. 12, pp. 667–674, 2003.
- [2] I. K. Macdonald, M. Harkiolaki, L. Hunt et al., "MHC class I bound to an immunodominant *Theileria parva* epitope demonstrates unconventional presentation to T cell receptors," *PLoS Pathogens*, vol. 6, no. 10, Article ID e1001149, 2010.
- [3] B. Korber, M. LaBute, and K. Yusim, "Immunoinformatics comes of age," *PLoS Computational Biology*, vol. 2, no. 6, Article ID e71, 2006.
- [4] I. A. Doytchinova and D. R. Flower, "Quantitative approaches to computational vaccinology," *Immunology and Cell Biology*, vol. 80, no. 3, pp. 270–279, 2002.
- [5] D. R. Flower, "Databases and data mining for computational vaccinology," *Current Opinion in Drug Discovery and Development*, vol. 6, no. 3, pp. 396–400, 2003.
- [6] S. Vivona, J. L. Gardy, S. Ramachandran et al., "Computer-aided biotechnology: from immuno-informatics to reverse vaccinology," *Trends in Biotechnology*, vol. 26, no. 4, pp. 190–200, 2008.
- [7] A. Chatzikyriakidou, P. V. Voulgari, and A. A. Drosos, "What is the role of HLA-B27 in spondyloarthropathies?" *Autoimmunity Reviews*, vol. 10, no. 8, pp. 464–468, 2011.
- [8] M. Marcilla and J. A. López de Castro, "Peptides: the cornerstone of HLA-B27 biology and pathogenetic role in spondyloarthritis," *Tissue Antigens*, vol. 71, no. 6, pp. 495–506, 2008.
- [9] J. A. Lopez de Castro, I. Alvarez, M. Marcilla et al., "HLA-B27: a registry of constitutive peptide ligands," *Tissue Antigens*, vol. 63, no. 5, pp. 424–445, 2004.
- [10] J. A. López de Castro, "The HLA-B27 peptidome: building on the cornerstone," *Arthritis and Rheumatism*, vol. 62, no. 2, pp. 316–319, 2010.
- [11] J. A. López de Castro, "HLA-B27 and the pathogenesis of spondyloarthropathies," *Immunology Letters*, vol. 108, no. 1, pp. 27–33, 2007.
- [12] K. Falk, O. Rotzschke, S. Stevanovic, G. Jung, and H.-G. Rammensee, "Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules," *Nature*, vol. 351, no. 6324, pp. 290–296, 1991.
- [13] J. Ruppert, J. Sidney, E. Celis, R. T. Kubo, H. M. Grey, and A. Sette, "Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules," *Cell*, vol. 74, no. 5, pp. 929–937, 1993.
- [14] I. A. Doytchinova, V. A. Walshe, N. A. Jones, S. E. Gloster, P. Borrow, and D. R. Flower, "Coupling *in silico* and *in vitro* analysis of peptide-MHC binding: a bioinformatic approach enabling prediction of superbinding peptides and anchorless epitopes," *The Journal of Immunology*, vol. 172, no. 12, pp. 7495–7502, 2004.
- [15] I. A. Doytchinova, V. Walshe, P. Borrow, and D. R. Flower, "Towards the chemometric dissection of peptide–HLA-A\*0201 binding affinity: comparison of local and global QSAR models," *Journal of Computer-Aided Molecular Design*, vol. 19, no. 3, pp. 203–212, 2005.
- [16] P. Guan, I. A. Doytchinova, V. A. Walshe, P. Borrow, and D. R. Flower, "Analysis of peptide-protein binding using amino acid descriptors: prediction and experimental verification for human histocompatibility complex HLA-A\*0201," *Journal of Medicinal Chemistry*, vol. 48, no. 23, pp. 7418–7425, 2005.
- [17] V. A. Walshe, C. K. Hattotuagama, I. A. Doytchinova et al., "Integrating *in silico* and *in vitro* analysis of peptide binding affinity to HLA-Cw\*0102: a bioinformatic approach to the prediction of new epitopes," *PLoS ONE*, vol. 4, no. 11, Article ID e8095, 2009.
- [18] C. Lundegaard, K. Lamberth, M. Harndahl, S. Buus, O. Lund, and M. Nielsen, "NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11," *Nucleic Acids Research*, vol. 36, supplement 2, pp. W509–W512, 2008.
- [19] B. Peters and A. Sette, "Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method," *BMC Bioinformatics*, vol. 6, article 132, 2005.
- [20] J. Sidney, E. Assarsson, C. Moore et al., "Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries," *Immunome Research*, vol. 4, article 2, 2008.
- [21] Y. Kim, J. Ponomarenko, Z. Zhu et al., "Immune epitope database analysis resource," *Nucleic Acids Research*, vol. 40, no. W1, pp. W525–W530, 2012.



**Hindawi**  
Submit your manuscripts at  
<http://www.hindawi.com>

