

Reviewed Preprint

v1 • June 18, 2026

Not revised

Reviewed Preprint

v2 • August 26, 2026

Revised by authors

✉ For correspondence:

susan.lea@stjude.org

nternette001@dundee.ac.uk

jim.kaufman@ed.ac.uk

*, †, ‡, § **Author notes:** See [page 46](#)

Funding: See [page 46](#)

Reviewing editor: Jungsan Sohn,
Johns Hopkins University School of
Medicine, United States

© 2026, Harrison et al. This article is
distributed under the terms of the

[Creative Commons Attribution](#)

[License](#), which permits unrestricted
use and redistribution provided that
the original author and source are
credited.

Simplifying principles underlying the highly complex peptide motif of the promiscuous chicken MHC class I molecule, BF2*21:01

Michael Harrison¹, Paul E Chappell², Samer Halabi^{3,*}, Maria Danysz³, Enock M Mararo³, Łukasz Magiera^{1,†}, Clemens Hermann¹, Michael J Deery⁴, Kathryn S Lilley⁴, Hans-Joachim Wallny^{5,‡}, David W Avila⁵, William Mwangi^{6,7,§}, Venugopal Nair^{6,7,8}, Susan M Lea^{2,9}✉, Nicola Ternette^{10,11,12}✉, Jim Kaufman^{1,3,5,6,13,14}✉

¹University of Cambridge, Department of Pathology, Cambridge, United Kingdom • ²University of Oxford, Sir William Dunn School of Pathology, Oxford, United Kingdom • ³University of Edinburgh, Institute for Immunology and Infection Research, Edinburgh, United Kingdom • ⁴University of Cambridge, Department of Biochemistry, Cambridge Centre for Proteomics, Cambridge, United Kingdom • ⁵Basel Institute for Immunology, Basel, Switzerland • ⁶Institute for Animal Health, Compton, United Kingdom • ⁷Pirbright Institute, Pirbright, United Kingdom • ⁸University of Oxford, Department of Biology, Oxford, United Kingdom • ⁹St Jude Children's Research Hospital, Department of Structural Biology, Memphis, United States • ¹⁰University of Oxford, Centre for Immuno-Oncology, Headington, United Kingdom • ¹¹University of Oxford, Jenner Institute, Headington, United Kingdom • ¹²University of Dundee, School of Life Sciences, Dundee, United Kingdom • ¹³University of Cambridge, Department of Veterinary Medicine, Cambridge, United Kingdom • ¹⁴Yale University, Department of Immunobiology, New Haven, United States

eLife Assessment

This **important** study investigates the peptide-binding principles by promiscuous chicken MHC molecules. The data from crystallography, mass-spec, and modeling are **compelling**. This paper will be of broad interest to immunologists and those interested in vaccine development.

<https://doi.org/10.7554/eLife.111098.2.sa3>

Abstract

In chickens and humans, classical class I molecules of the major histocompatibility complex (MHC) have a spectrum of correlated properties, including cell surface expression and peptide repertoire. Understanding the impact of these promiscuous generalists and fastidious specialists is of considerable interest. The promiscuous BF2 molecule from the chicken B21 haplotype, BF2*21:01, binds a wide range of peptides by remodelling the peptide-binding site, allowing co-variation of the anchor residues at peptide positions P₂ and P_{C-2}, and binding of an anchor residue at P_C. Using in vitro refolding assays with peptides and peptide libraries, determining thermostability and crystal structures, and analysing a chicken B21 cell line by immunopeptidomics, we found that BF2*21:01 will accommodate many possible combinations at P₂ and P_{C-2}, as well as several hydrophobic amino acids at P_C. However, marked preferences for particular peptide lengths, particular amino acids at the three anchor residues, and particular combinations of amino acids at P₂ and P_{C-2} as well as at P₃ and P_{C-3} together lead to high frequencies of major peptides while still allowing the possibility of presenting a wide peptide repertoire. These simplifying principles may eventually allow predictions of pathogen peptides with stable binding for this iconic promiscuous class I molecule, as well as providing the data for more sophisticated peptide prediction methods.

Introduction

The major histocompatibility complex (MHC) is a genetic region defined by the presence of a few highly polymorphic genes encoding classical class I and class II molecules, originally identified as transplantation antigens but now known to play crucial roles in the immune response to pathogens and tumours (Djaoud and Parham, 2020). The human MHC is an enormous genomic region with much recombination among hundreds of genes with a vast array of functions, with multigene families of classical class I and class II genes, strongly associated with autoimmunity and allergies but not so strongly with infectious pathogens (Trowsdale and Knight, 2013). In contrast, the chicken MHC is much smaller and simpler, with single dominantly-expressed class I and class II genes, and strong associations of MHC haplotypes with resistance and susceptibility to certain economically-important pathogens (Kaufman 2018; Miller et al 2016; Tregaskes and Kaufman, 2021).

In trying to understand the strong associations of the chicken MHC haplotypes with infectious disease, we discovered a hierarchy of alleles of the dominantly-expressed class I molecule, BF2. The original insight came from flow cytometry: erythrocytes from the B21 haplotype conferring resistance to the oncogenic herpesvirus that causes Marek's disease have ten-fold less class I on the cell surface than from the susceptible B4, B15, B12 and B19 haplotypes (Kaufman et al., 1995). Since then, a suite of properties was found to correlate with this hierarchy, the most important being the peptide repertoire, defined as the range of different peptide sequences bound. An inverse correlation between the level of cell surface expression and the breadth of peptide binding was found, with less well-expressed alleles that have wide peptide repertoires correlating with resistance to Marek's disease and other economically-important pathogens (Chappell et al., 2015; Kaufman, 2015; Kaufman, 2018; Tregaskes and Kaufman, 2021).

A similar inverse correlation between predicted peptide repertoire and cell surface expression was found for some human class I molecules, but a disease association that was the opposite: well-expressed "elite controller" alleles with narrow peptide repertoires correlated with slow progression from HIV infection to AIDS (Chappell et al., 2015). This finding led to the hypothesis of generalist and specialist class I alleles, with low-expressing promiscuous alleles generally protecting from many pathogens while high-expressing fastidious alleles bound special peptides to protect against particular pathogens (Chappell et al., 2015; Kaufman, 2018). A hierarchy of tapasin-dependence and ease of in vitro refolding for certain human class I alleles (Rizvi et al., 2014) was found to correlate with cell surface expression and predicted peptide repertoire (Kaufman, 2015; Kaufman, 2018; Tregaskes and Kaufman, 2021) and eventually shown to associate with slow progression once elite controllers were removed from analysis (Bashirova et al., 2020), thus supporting the generalist-specialist hypothesis. In addition, a hierarchy of human class II alleles based on predicted peptide repertoire was reported (Manczinger et al., 2019).

There are presumably various biochemical mechanisms leading to this hierarchy of class I alleles, including specificity of peptide transport by TAPs in chickens (Tregaskes et al., 2016; Walker et al., 2011), dependence of class I alleles on tapasin (and perhaps TAPBPR) in humans (Bashirova et al., 2020; Kaufman, 2015; Kaufman, 2018; Rizvi et al., 2014), and specificity of peptide binding by the class I alleles themselves (Chappell et al., 2015; Han et al., 2023; Koch et al., 2008; Paul et al., 2013; Shaw et al., 2007; Wallny et al., 2006; Xiao et al., 2018; Zhang et al., 2012). Several modes of promiscuous binding have been reported for dominantly-expressed chicken class I molecules (Halabi and Kaufman, 2022).

Despite requiring three anchor residues, BF2*21:01 achieves a wide peptide repertoire by remodelling the peptide-binding site in a previously unprecedented manner that involves co-variation between the amino acids at positions P₂ and P_{C-2}, as shown by structures bearing six peptides with very divergent sequences (Chappell et al., 2015; Koch et al., 2008). Since this iconic chicken class I molecule is so frequent across the world and so important for resistance to economically-important viral diseases (Kaufman, 2018), we wished to better understand the rules by which peptides bind to BF2*21:01, starting with refolding the class I molecule with

different peptides and peptide libraries, determining thermostabilities, determining structures and creating models to illustrate some of the results, and finishing with immunopeptidomics of class I molecules from a B21 cell line.

Results

The class I molecule(s) from the B21 haplotype bind a wide variety of peptides

Gas phase sequencing of peptides eluted from B21 class I molecules and separated by HPLC gave a wide range of sizes and sequences with no obvious anchor residues (Fig. 1 [↗](#)), in stark contrast to the class I molecules from B4, B12, B15 and B19 which gave mostly octamers with clear peptide motifs (Kaufman et al., 1995 [↗](#); Wallny et al., 2006 [↗](#)). The heterogeneity in length of peptides from blood and spleen cells is likely to have been due to proteolysis, since the single preparation from a B21 chicken cell line gave only 10mers and 11mers, including TNPEKVFYL from which the octamer PESKVFYL from blood samples (that failed to refold in vitro) was apparently derived. The wide range of sequences were eventually understood to mean that BF2*21:01 molecules would accommodate a variety of anchor residues at three positions, with co-variation between the anchor residues at P₂ and P_{C-2} (position 2 and the position two before the C-terminus of the peptide), as found by refolding and crystallography (Chappell et al., 2015 [↗](#); Koch et al., 2008).

To better understand the requirements of peptide binding to BF2*21:01, refolding in vitro followed by size exclusion chromatography (hereafter called “assembly”) was carried out with individual peptides. Four outcomes were routinely found: no monomer peak reflecting no assembly, a relatively sharp monomer peak reflecting a stable binding, a broader monomer peak that reflected monomers falling apart during chromatography and a sharp peak at the position of BF2*21:01 heavy chain alone that presumably reflected monomers that fell apart before or during chromatography (Fig. 2 [↗](#)).

Starting with two peptides for which the first crystal structures for BF2*21:01 were determined (Koch et al., 2008), assembly assays with single peptides substituted at positions P₂, P_{C-2} and P_C showed that some (combinations of) amino acids were tolerated while others were not (Fig. 3 [↗](#)). For the 11mer GHAEYGAETL, the basic amino acid His at P₂ allowed both acidic Asp and Glu at P_{C-2} but the hydrophobic Leu at P₂ gave unstable binding. However, substituting basic Lys and Arg at P₂ only allowed Asp at P_{C-2}. The 10mer REVDEQLLSV allowed many hydrophobic amino acids (except Ala) at P_{C-2}, but substituting acidic Asp at P₂ allowed no stable peptide binding with the same amino acids at P_{C-2}. Indeed, both the 11mer and the 10mer peptides with substitutions of hydrophobic amino acids at P₂ failed to bind in a stable way. The 11mer substituted with a Glu at P₂ alone or with Leu at P_{C-2} (as in the 10mer) only bound well when P₃ was changed to Val (as in the 10mer). In contrast, the 10mer tolerated His at P₂ and Glu at P_{C-2} (as in the 11mer), except with Ala at both P₃ and P_{C-3} (as in the 10mer). Substitution at P_C in both the 10mer and 11mer showed that Leu, Phe and Val but not Ala or Trp were tolerated. Assembly assays with derivatives of the 11mer show that the 10mer and 11mer bound well, but 9mers and below were not as stable (Fig. 4 [↗](#)). Overall, the rules for binding seemed complex.

Some of these diverse results could be understood by structures and models of the peptides bound to BF2*21:01. Structures with amino acid substitutions in the 11mer peptide GHAEYGAETL bind with Asp at P_{C-2} and either His or Arg at P₂ (5AD0 and 5ADZ), comparable to the original structure (3BEV) (Fig. 5A [↗](#)). Modelling the substituted peptide with Arg at P₂ and Glu at P_{C-2} shows a steric clash that can explain why this peptide did not refold with BF2*21:01 (Fig. 5A [↗](#)). At P_C, substitution of a Trp instead of the original Leu in this 11mer peptide leads to steric clashes with pocket F of BF2*21:01 (Fig. 5B [↗](#)).

The thermostability of BF2*21:01 molecules refolded with various peptides confirms and extends these findings (Fig. 6 [↗](#)). From the peptides eluted from B21 cells, one 11mer and eight 10mers (of which the structures have been determined for five, Chappell et al., 2015 [↗](#); Koch et al., 2008) were tested. The 10mers REVDEQLLSV and TNPEKVFYL were the most stable and the 10mer RSINGLDALKL and the 11mer GHPDWYYGETW were the least stable. The 11mer GHAEYGAETL (a

Basel H.B21 line blood	
STA	GHAEYGAETL hemoglobin alpha-A chain (11/11) NP_001004376
TA	GHDPWYGETW cilia & flagella assoc protein (5/11) XP_040518076
I	ELHTRLRYIRTA MHC class I (BF) heavy chain (11/11) NP_001026509
	PEPDARYERTTA smoothin (7/11) XP_040540967
	GHPEEYAAET van Willebrand factor A protein 8 (7/10) XP_04051448
	RHPEEYSMGL WD repeat coiled coil protein (6/10) XP_040524133
	RHPEEYARI van Willebrand factor A protein 8 (7/10) XP_04051448
	RLVDLRYIH dystonin (5/9) XP_046770657
I	PEKVFY (part of TNPEKVFYL, below)
	TKVFY
Basel H.B21 line blood	
	TNDLVSEYQYQY tubulin beta chain (11/12) NP_990776
	MLEIVFLQIGQ tubulin beta chain (9/11) NP_990776
	PEFTL
Basel H.B21 line blood	
STA	GHAEYGAETL hemoglobin alpha-A chain (11/11) NP_001004376
	XIQMTQSPASL voltage-gated Ca++ channel (8/11) XP_040551278
STA	REVDEQLLEV tubulin beta chain (10/10) NP_990776
	VEPDVNYEE 26S proteasome regulatory (8/9) NP_001026361
	SFTELDYV
	SFNKXES (part of MLTGNYTLSTL, below)
	VHNDP
I	SDL
I	GEL
I	IQM (part of xIQMTQSPASL, above)
I	PEF (part of PEFTL, above)
Compton N line spleen	
STA	GHAEYGAETL hemoglobin alpha-A chain (11/11) NP_001004376
STAI	TAGQEDYDRL transforming protein RhoA (10/10) XP_046781954
	TVLSTKLEFL core histone macro H2A (10/10) XP_990338
T	QHNYGILESF *uncharac LOC101747454 (9/10) NP_001305924 (also BLB)
TA	RHNYGILESF *uncharac LOC101747454 (9/10) NP_001305924 (also BLB)
	RSINLDALK (part of RSINLDALKL, below)
	RSINIDAL (part of RSINLDALKL, below)
	VSVALLFY
Compton TG21 REV cell line	
S	TNDNVSEYQYQY tubulin beta chain (10/11) NP_990776
	SNYEEIYQVLK Niban 1 (8/11) XP_046778561
	MLTGNYTLSTL NADH dehydrogenase (10/11) YP_009555261
T	RSINLDALKL bax inhibitor (8/10) XP_040510747
	RHPxLKKTY puratrophin 1 (7/9) XP_040536880
Compton N line spleen	
	KEIDSVKYLEK ras-related C3 botulinum substrate (10/11) XP_001188381
STAI	TNPEKVFY 14-3-3 protein theta (10/10) NP_001006415
ST	TAGQSNVDRL transforming protein RhoA (8/10) XP_046781954
STA	YELDEKFDRL septin 6 (8/10) XP_015133785
	YDPDEYEQ spectrin beta chain (6/9) XP_040557384

Figure 1. Peptides isolated from class I molecules of chicken blood, spleen and a B21 cell line separated by HPLC with single peaks subjected to gas phase sequencing show a range of lengths and sequences.

Methods are detailed in Chappell et al., 2015 [\[1\]](#); Koch et al., 2008; Wallny et al., 2006 [\[2\]](#). The peptides are grouped by the six experiments with locations and cell sources indicated. Sequences are in single letter amino acid code, small letter indicates dominant amino acid in a position, and x indicates ambiguous call. The amino acids are coloured throughout this paper according to their primary characteristic (with the understanding that some amino acids could be characterised in more than one way, eg. Y both hydrophobic and polar, K basic and hydrophobic, etc): red, acidic, D and E; blue, basic, H, K and R; green, polar, N, Q, S and T; black, hydrophobic, A, C, F, G, I, L, M, P, V, W and Y). Designations on the left indicate synthetic peptides used for structures (S), thermostability assays (T) or assembly assays (A), or found as natural peptide by immunopeptidomics (I). Chicken genes from which a (partial) peptide sequence could be identified are presented, with the number of amino acids positions matching compared to total length in brackets.

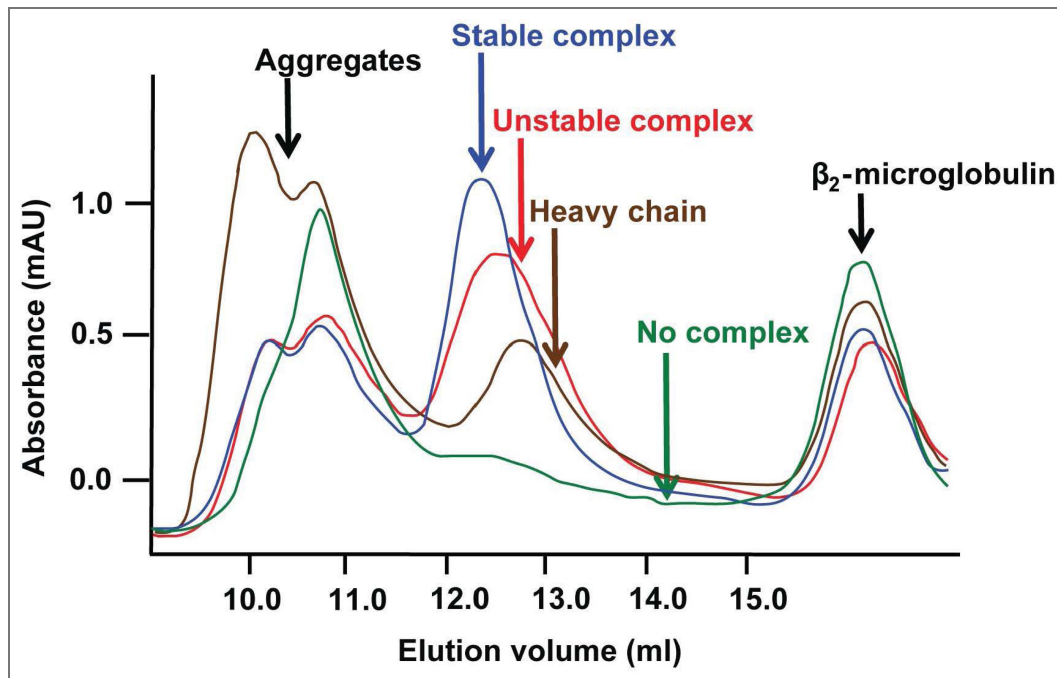


Figure 2. Representative SEC trace for assembly assays show four typical outcomes for BF2*21:01.

After refolding of bacterially-expressed heavy chain and β_2 -microglobulin with synthetic peptides according to published methods (Chappell et al., 2015; Koch et al., 2008), samples were spun in a cooled microfuge at full speed for 30 minutes, the supernatant was put through a 0.2 microfilter and then the flow-through was loaded on a SEC column (in this case, a Superdex 200 10/300 GL column) attached to an AKTA FPLC instrument. In every run, peaks at and slightly after the exclusion volume represent aggregated material, and a peak before the inclusion body represented refolded β_2 -microglobulin. In between, a sharp peak represented stably refolded monomers of heavy chain, β_2 -microglobulin and peptide (peptide with blue trace), a broader peak beginning at the same place as the monomer peak presumably represented refolded monomers that were dissociating during chromatography (red), a later broad was at the same position as heavy chain refolded alone (brown) and for some peptides no peak in between the aggregate peaks and the β_2 -microglobulin peak, representing no successful refolding (green). The x-axis represents elution volume in ml, while the y-axis represents OD₂₁₅ in mAU.

11mer			10mer		
P316	<u>G</u> <u>H</u> AEEYGA <u>E</u> <u>T</u> <u>L</u>	+	P330	<u>R</u> <u>E</u> VDEQL <u>L</u> <u>S</u> <u>V</u>	+
P405	<u>G</u> <u>H</u> AEEYGA <u>D</u> <u>T</u> <u>L</u>	+	P439	<u>R</u> <u>E</u> VDEQL <u>A</u> <u>S</u> <u>V</u>	h
P406	<u>G</u> <u>H</u> AEEYGA <u>L</u> <u>T</u> <u>L</u>	b	P440	<u>R</u> <u>E</u> VDEQL <u>V</u> <u>S</u> <u>V</u>	+
P407	<u>G</u> <u>K</u> AEEYGA <u>E</u> <u>T</u> <u>L</u>	0	P441	<u>R</u> <u>E</u> VDEQL <u>I</u> <u>S</u> <u>V</u>	+
P408	<u>G</u> <u>K</u> AEEYGA <u>D</u> <u>T</u> <u>L</u>	+	P442	<u>R</u> <u>E</u> VDEQL <u>M</u> <u>S</u> <u>V</u>	+
P409	<u>G</u> <u>K</u> AEEYGA <u>L</u> <u>T</u> <u>L</u>	0	P443	<u>R</u> <u>E</u> VDEQL <u>F</u> <u>S</u> <u>V</u>	+
P410	<u>G</u> <u>R</u> AEEYGA <u>E</u> <u>T</u> <u>L</u>	b	P444	<u>R</u> <u>D</u> VDEQL <u>L</u> <u>S</u> <u>V</u>	b
P411	<u>G</u> <u>R</u> AEEYGA <u>D</u> <u>T</u> <u>L</u>	+	P445	<u>R</u> <u>D</u> VDEQL <u>A</u> <u>S</u> <u>V</u>	h
P412	<u>G</u> <u>R</u> AEEYGA <u>L</u> <u>T</u> <u>L</u>	h	P446	<u>R</u> <u>D</u> VDEQL <u>V</u> <u>S</u> <u>V</u>	h
P422	<u>G</u> <u>H</u> AEEYGA <u>E</u> <u>T</u> <u>A</u>	h	P447	<u>R</u> <u>D</u> VDEQL <u>I</u> <u>S</u> <u>V</u>	h
P423	<u>G</u> <u>H</u> AEEYGA <u>E</u> <u>T</u> <u>F</u>	+	P448	<u>R</u> <u>D</u> VDEQL <u>M</u> <u>S</u> <u>V</u>	h
P424	<u>G</u> <u>H</u> AEEYGA <u>E</u> <u>T</u> <u>W</u>	h	P449	<u>R</u> <u>D</u> VDEQL <u>F</u> <u>S</u> <u>V</u>	b
P419	<u>G</u> <u>V</u> AEEYGA <u>L</u> <u>T</u> <u>L</u>	h	P435	<u>R</u> <u>E</u> VDEQL <u>L</u> <u>S</u> <u>A</u>	b
P420	<u>G</u> <u>M</u> AEEYGA <u>L</u> <u>T</u> <u>L</u>	h	P436	<u>R</u> <u>E</u> VDEQL <u>L</u> <u>S</u> <u>F</u>	+
P421	<u>G</u> <u>M</u> AEEYGA <u>V</u> <u>T</u> <u>L</u>	h	P438	<u>R</u> <u>E</u> VDEQL <u>L</u> <u>S</u> <u>V</u>	+
P413	<u>G</u> <u>E</u> AEEYGA <u>E</u> <u>T</u> <u>L</u>	h	P437	<u>R</u> <u>E</u> VDEQL <u>L</u> <u>S</u> <u>W</u>	h
P414	<u>G</u> <u>E</u> AEEYGA <u>L</u> <u>T</u> <u>L</u>	h	P430	<u>R</u> <u>A</u> VDEQL <u>L</u> <u>S</u> <u>V</u>	h
P415	<u>G</u> <u>E</u> VEEYGA <u>L</u> <u>T</u> <u>L</u>	+	P431	<u>R</u> <u>V</u> VDEQL <u>L</u> <u>S</u> <u>V</u>	0
P416	<u>G</u> <u>E</u> AEEY <u>G</u> <u>V</u> <u>L</u> <u>T</u> <u>L</u>	h	P432	<u>R</u> <u>I</u> VDEQL <u>L</u> <u>S</u> <u>V</u>	h
P417	<u>G</u> <u>E</u> VEEY <u>G</u> <u>V</u> <u>L</u> <u>T</u> <u>L</u>	b	P433	<u>R</u> <u>M</u> VDEQL <u>L</u> <u>S</u> <u>V</u>	h
P418	<u>G</u> <u>E</u> <u>A</u> <u>E</u> .Y <u>G</u> <u>V</u> <u>L</u> <u>T</u> <u>L</u>	h	P434	<u>R</u> <u>I</u> VDEQL <u>L</u> <u>S</u> <u>V</u>	h
P425	<u>R</u> <u>H</u> VDEQL <u>E</u> <u>S</u> <u>V</u>	+	P425	<u>R</u> <u>H</u> VDEQL <u>E</u> <u>S</u> <u>V</u>	+
P426	<u>R</u> <u>H</u> ADEQL <u>E</u> <u>S</u> <u>V</u>	+	P426	<u>R</u> <u>H</u> ADEQL <u>E</u> <u>S</u> <u>V</u>	+
P427	<u>R</u> <u>H</u> VDEQL <u>A</u> <u>E</u> <u>S</u> <u>V</u>	+	P427	<u>R</u> <u>H</u> VDEQL <u>A</u> <u>E</u> <u>S</u> <u>V</u>	+
P428	<u>R</u> <u>H</u> ADEQL <u>A</u> <u>E</u> <u>S</u> <u>V</u>	b	P428	<u>R</u> <u>H</u> ADEQL <u>A</u> <u>E</u> <u>S</u> <u>V</u>	b
P429	<u>R</u> <u>H</u> ADEQL <u>A</u> <u>E</u> <u>S</u> <u>V</u>	+	P429	<u>R</u> <u>H</u> ADEQL <u>A</u> <u>E</u> <u>S</u> <u>V</u>	+

Figure 3. Assembly of BF2*21:01 molecules by refolding in vitro is markedly affected by single amino acid differences between peptides.

Bacterial-expressed heavy chain and β_2 -microglobulin were refolded in vitro with single synthetic peptides (in-house identification of peptide batch by P and number), shown as single letter code with anchor residues underlined and coloured (red, acidic; blue, basic; black hydrophobic), changes from original peptide bolded, and a dot to indicate a residue skipped. Result after analysis by SEC given as +, sharp peak at position of monomer; b, broad peak starting at position of monomer; h, sharp peak at position of heavy chain alone; 0, no peak in between aggregated material at the exclusion volume of the column and the peak representing refolded β_2 -microglobulin. Methods as in Fig. 2.

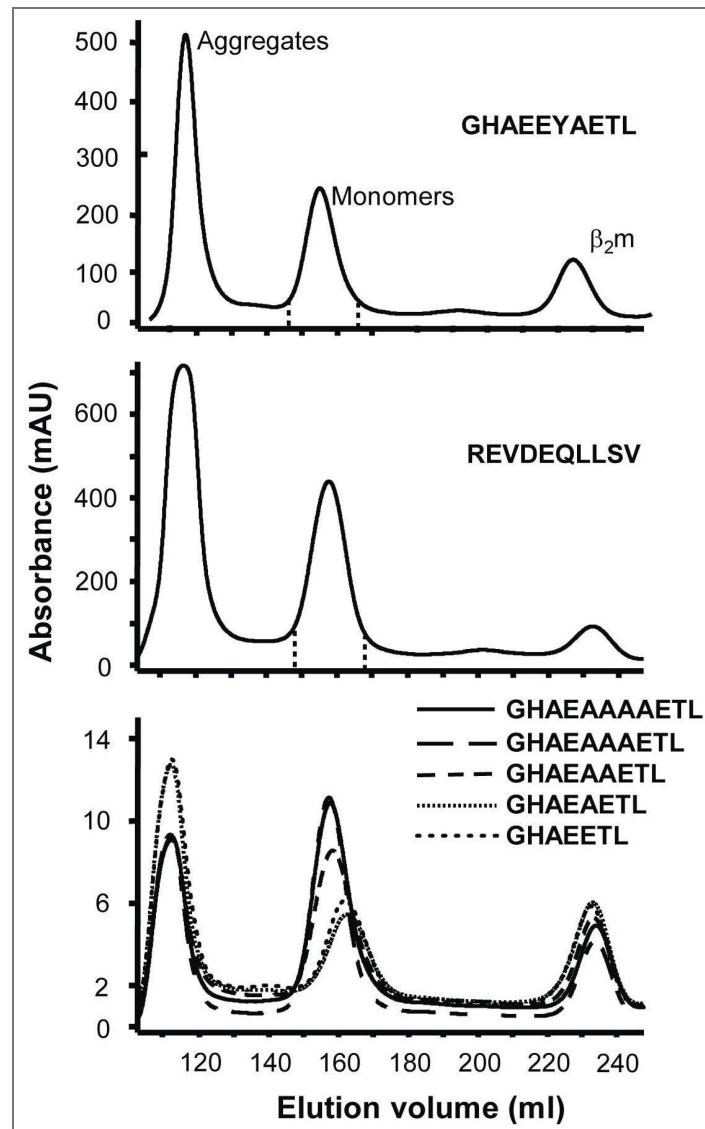


Figure 4. The original 10mer (REVDEQLLSV) and 11mer (GHAEYYAETL) peptides refold with BF2*21:01 to give stable monomers as do 11mer and 10mer derivative peptides, but 9mer, 8mer or 7mer peptides give heavy chain only peaks.

Assembly assays for representative 10mer and 11mer peptides give peaks for stable monomers of BF2*21:01, but shorter peptides based on the 11mer do not. Single synthetic peptides were refolded with bacterial-expressed heavy chain and β_2 -microglobulin with synthetic peptides, prepared for SEC as described in the legend for Fig. 2, and analysed in this case with a HiLoad 26/60 Superdex 200 column attached to an AKTA FPLC instrument. The peptides 11mer GHAEYYAETL (top panel), 10mer REVDEQLLSV (middle panel), 11mer GHAEAAAETL and 10mer GHAEAAAETL (bottom panel) gave sharp monomer peaks, while the 9mer GHAEAAETL, 8mer GHAEAETL and 7mer GHAEETL gave a delayed broad peak indicative of unstable binding or heavy chain.

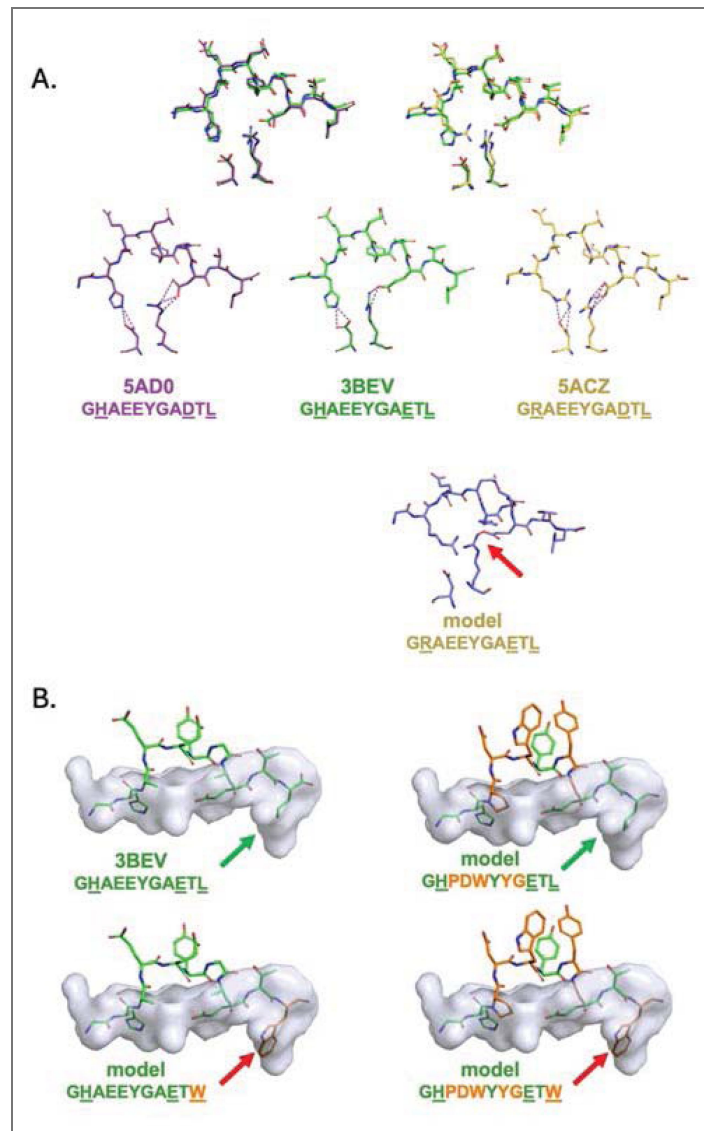


Figure 5. Crystal structures of BF2*21:01 with peptides that refold to give stable monomers and models of structures with peptides that either did not refold to give stable monomers or refolded to give monomers with low thermal stability.

A. Structures were determined for GHAEYGAETL (3BEV), GHAEYGAETL (5AD0) and GRAEYGAETL (5ACZ), which all refolded successful to give stable monomers, while GRAEYGAETL did not (see Fig. 3), all models of which showed steric clashes (one depicted, red arrow). B. Two 11mers refolded to give monomers, which were much more stable for GHAEYGAETL than GHAEYGAETW (see Fig. 6). Models of GHAEYGAETL and GHAEYGAETW show steric clashes (red arrows), but the crystal structure with GHAEYGAETL (3BEV) and a model of GHAEYGAETL do not (green arrows). Underlined positions indicate anchor residues; bold letters indicate amino acid substitutions, modelled positions in orange. Methods for refolding and analysis as in Fig. 2, modelling done as detailed in Materials and Methods, structure from Koch et al., 2008.

derivative of the original GHAEYGAETL) was of middling stability, but substitution of P₂ with Arg or Lys reduced the stability to the same as the 11mer GHPDWYYGETW. Similarly, the high stability of the 10mer REVDEQLLSV was reduced by substitution of P_{C-2} to Val, Ile or Met, or by the combination of His at P₂ and Glu at P_{C-2} (as in the 11mer GHAEYGAETL). The substitution of Phe at P_C lowered the stability of REVDEQLLSV only slightly, but the substitution of Leu at P_C in GHPDWYYGETW raised the stability significantly, consonant with models of the two peptides bound to BF2*21:01 (Fig. 5B). Derivatives of the 11mer, GHAEAA(A,AA)ETL, and of the 10mer, RHAEQAESV, showed that 10mers were more stable than either 9mers or 11mers (Fig. 6).

Assembly assays with peptide libraries show great variation in anchor and peptide backbone residues

To expand the range of amino acids examined, assembly assays followed by mass spectrometry were conducted with peptide libraries based on the 11mer and 10mer, starting with single-substitution libraries assessed by MALDI-TOF (Fig. 6). The libraries contained roughly equimolar amounts of all amino acids except Cys; given the molecular mass, Ile and Leu could not be distinguished, nor could Gln and Lys. Two concentrations of peptide were used, one at the usual 10-fold molarity over heavy chain and one at only 2-fold, with the idea that at 10-fold (but not 2-fold) a particularly high affinity peptide might be in sufficient concentration to compete out other binding peptides. In fact, there were few cases where the two concentrations gave much difference.

Using these single-substitution libraries based on the 11mer GHAEYGAETL and the 10mer REVDEQLLSV (Fig. 7), substitution at P₂ for the 11mer showed predominantly His with much lesser amounts of Arg, Ser, Gln/Lys and Asn, while for the 10mer both His and Glu were predominant with much lesser amounts of Asn, Asp, Ala and Ser. Substitution at P_{C-2} for the 10mers REVDEQLLSV and RDVDEQLLSV showed Phe followed by Ile/Leu, Met, Gln/Lys and Val. For both the 11mer and 10mer, substitutions at P_C showed Ile/Leu and Phe followed by Met and Val. Substitutions at P₃ and P_{C-3} showed a variety of patterns.

To begin to understand the co-variation between P₂ and P_{C-2}, double-substitution libraries were created for the 11mer GHAEYGAETL and the 10mer REVDEQLLSV with the assembly assays analysed by MALDI-TOF (Fig. 8). This approach has serious limitations, not being able to distinguish between combinations that gave the same peaks (such as Arg-Ser and Asp-Lys) and/or which position the amino acids were in (Asp at P₂ and Ser at P_{C-2}, or vice versa). A further issue was the temperature stability of the complexes, so after refolding at 4°C but before chromatography at room temperature, aliquots were incubated for 18 hours at 4°C (as for the refolding) and at 42°C (roughly the body temperature of a chicken), compared to the control with no incubation. The proportions of most combinations did not change significantly between the control and 42°C incubation, but some combinations increased in proportion (four out of 15 combinations for the 11mer, and four out of 37 combinations for the 10mer), suggesting that such combinations are more stable or have slower dissociation times than those combinations that reduced in proportion. Perhaps the most important finding was that far fewer combinations of P₂ and P_{C-2} gave rise to stable peptide binding for the 11mer than for the 10mer, suggesting that this 11mer binds less stably and thus tolerates fewer changes than the 10mer.

The examination of double-substitution libraries analysed by MALDI-TOF was expanded to other backbones (that is, sequences in which two positions were randomised): five other 11mer peptides (GHPDWYYGETW and four derivatives of the 11mer GHAEYGAETL) and five 10mers (all from the original peptides found by gas phase sequencing, including the original REVDEQLLSV). For those libraries substituted at P₂ and P_{C-2}, four 11mers gave between 16 and 24 combinations, while the six 10mers gave between 32 and 55 combinations (Figs. 9, 10), so again the 10mers seemed to tolerate more combinations. Two other 11mer libraries examined substitutions at P₃ and P_{C-3}, which allowed many more combinations (Fig. 9).

To get a more accurate understanding of the substitutions, six of these double-substitution libraries were analysed by LC-MS/MS, which allowed an exact identity and position of the amino acids in the combinations in peptides that refolded with BF2*21:01 (Fig. 11 – Source Data 1). The

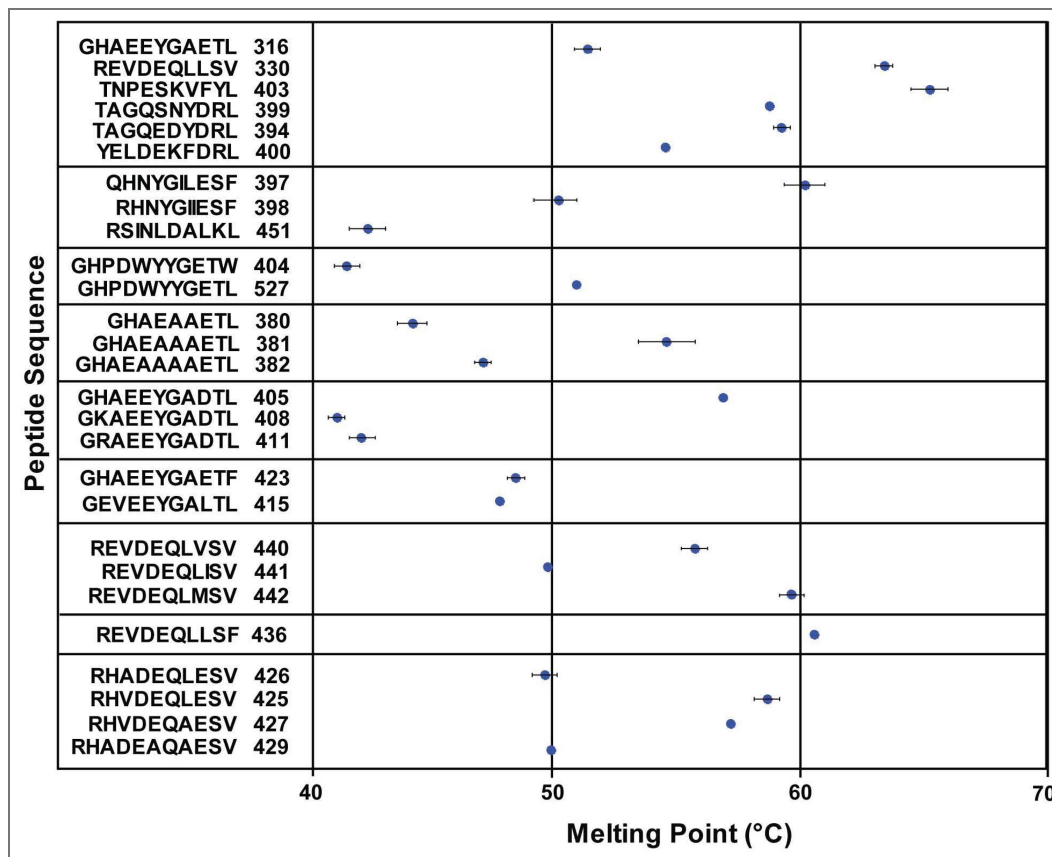


Figure 6. Peptide length and amino acids at peptide positions P_2 , P_3 , P_{C-3} , P_{C-2} and P_C affect stability of BF2*21:01 molecules.

Bacterial-expressed heavy chain and β_2 -microglobulin were refolded in vitro with various synthetic peptides and melting point measured as 50% uptake of a fluorescent dye that binds to hydrophobic regions, as described in Materials and Methods. Assays were performed in triplicate with error bars indicating standard errors. Peptide sequences are in single letter code with in-house identification of peptide batch by P and number.

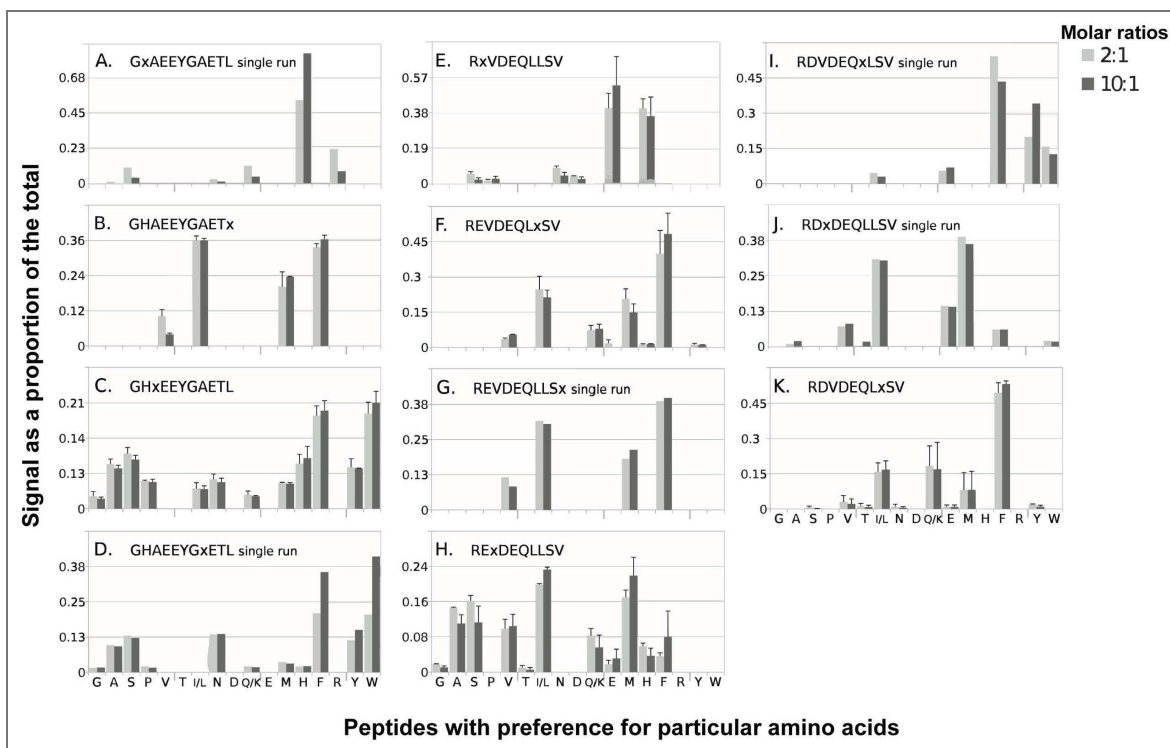


Figure 7. Assembly assays with single-substitution peptide libraries based on the 11mer GHAEYYAETL and the 10mer REVDEQLLSV show marked preferences for particular amino acids at P_2 , P_3 , P_{C-3} , P_{C-2} and P_C positions binding to BF2*21:01.

Bacterial-expressed heavy chain and β_2 -microglobulin were mixed with peptide libraries (19 amino acids, all except Cys, at one position) at two molar ratios (1:2:10 as usual, and 1:2:2 to control for strong binding peptides) and separated by SEC as described in the legend to Fig. 2; monomer peaks were collected, concentrated, treated with acid, lyophilised and analysed by MALDI-TOF. Bar charts show proportion of total signal (with error bars when present showing standard error around the mean for separate refolding experiments) on y-axis for mass/charge (m/z) positions corresponding to peptides with particular amino acids on the x-axis (single letter code; Ile and Leu were indistinguishable, as were Gln and Lys).

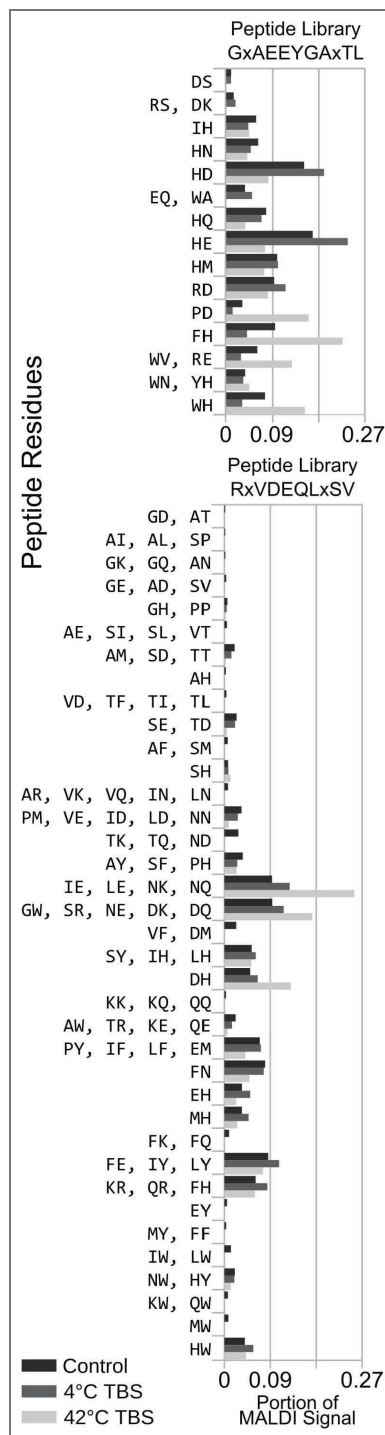


Figure 8. Assembly assays with double-substitution peptide libraries based on the 11mer GHAAEEYAETL (top panel) and the 10mer REVDEQLLSV (bottom panel) show marked preferences for particular amino acids combinations at P_2 and P_{C-2} positions binding to BF2*21:01, with more combinations found for the 10mer.

Bacterial-expressed heavy chain and β_2 -microglobulin were mixed with peptide libraries (19 amino acids, all except Cys, at two positions indicated by x), separated immediately by SEC as described in the legend to Fig. 2, or incubated at 4°C or 42°C for 18 hours before SEC; monomer peaks were collected, concentrated, treated with acid, lyophilised and analysed by MALDI-TOF. Bar charts show proportion of total signal (with error bars when present showing standard error around the mean for separate refolding experiments) on x-axis for mass/charge (m/z) positions corresponding to peptides with particular amino acid combinations on the y-axis (single letter code; Ile and Leu, Gln and Lys, and amino acids present in P_2 versus P_{C-2} positions were indistinguishable).

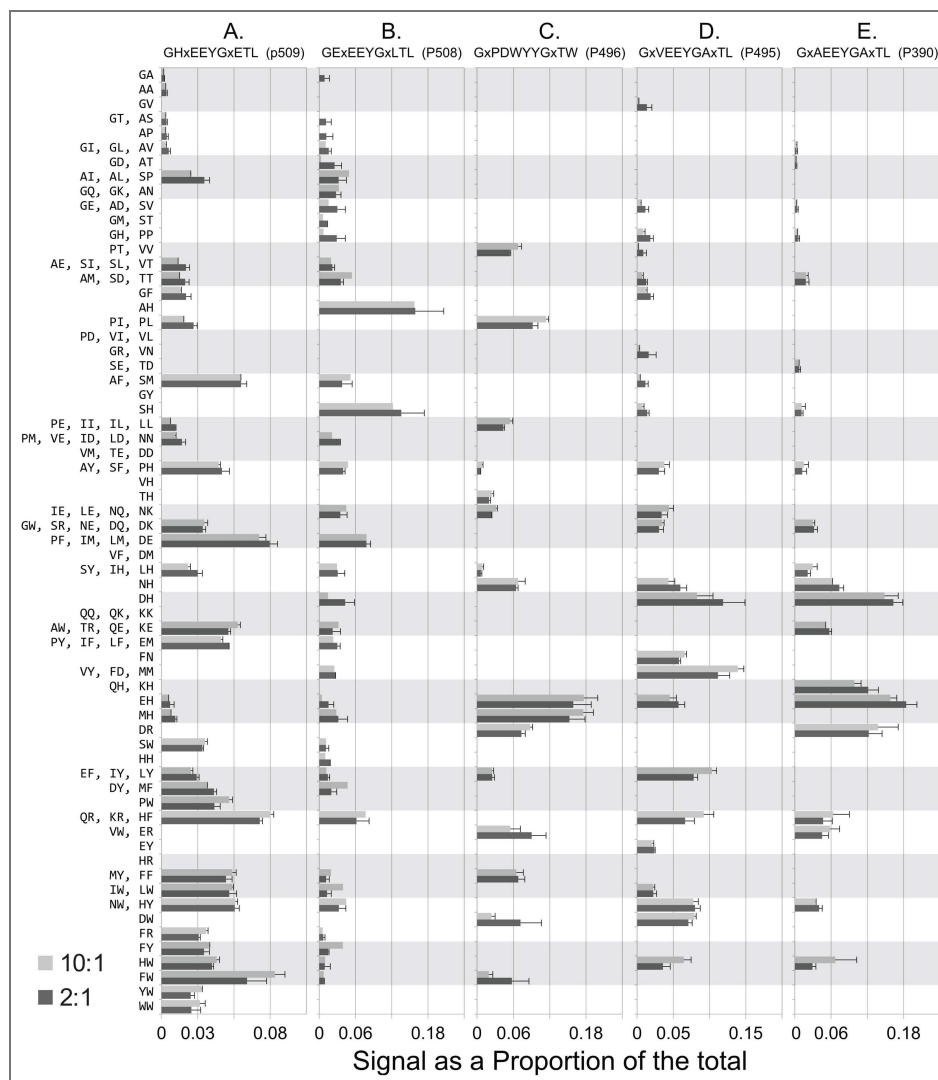


Figure 9.

Panels A-E. Assembly assays with double-substitution peptide libraries based on five 11mer peptides (GHPDWYYGETW and four derivatives of the 11mer GHAEYGAETL) and five 10mers (all from the original peptides found by gas phase sequencing, including the original REVDEQLLSV) show different preferences for particular amino acid combinations at P₂, P₃, P_{C-3} and P_{C-2} positions binding to BF2*21:01, with more combinations found for 10mer peptides. Bacterial-expressed heavy chain and β_2 -microglobulin were mixed with peptide libraries (19 amino acids, all except Cys, at one position, indicated by x in the peptide sequence) at two molar ratios (1:2:10 as usual, and 1:2:2 to control for strong binding peptides) and separated by SEC as described in the legend to Fig. 2 and separated immediately by SEC; monomer peaks were collected, concentrated, treated with acid, lyophilised and analysed by MALDI-TOF. Bar charts show proportion of total signal (with error bars when present showing standard error around the mean for separate refolding experiments) on x-axis for mass/charge (m/z) positions corresponding to peptides with particular amino acid combinations on the y-axis (single letter code; Ile and Leu, Gln and Lys, and amino acids present in P₂ versus P_{C-2} positions were indistinguishable). Panels: A. GHxEEYGxETL, B. GExEEYGxLTL, C. GxPDWYYGxTW, D. GxVEEYGxTL, E. GxAEEYGxTL.

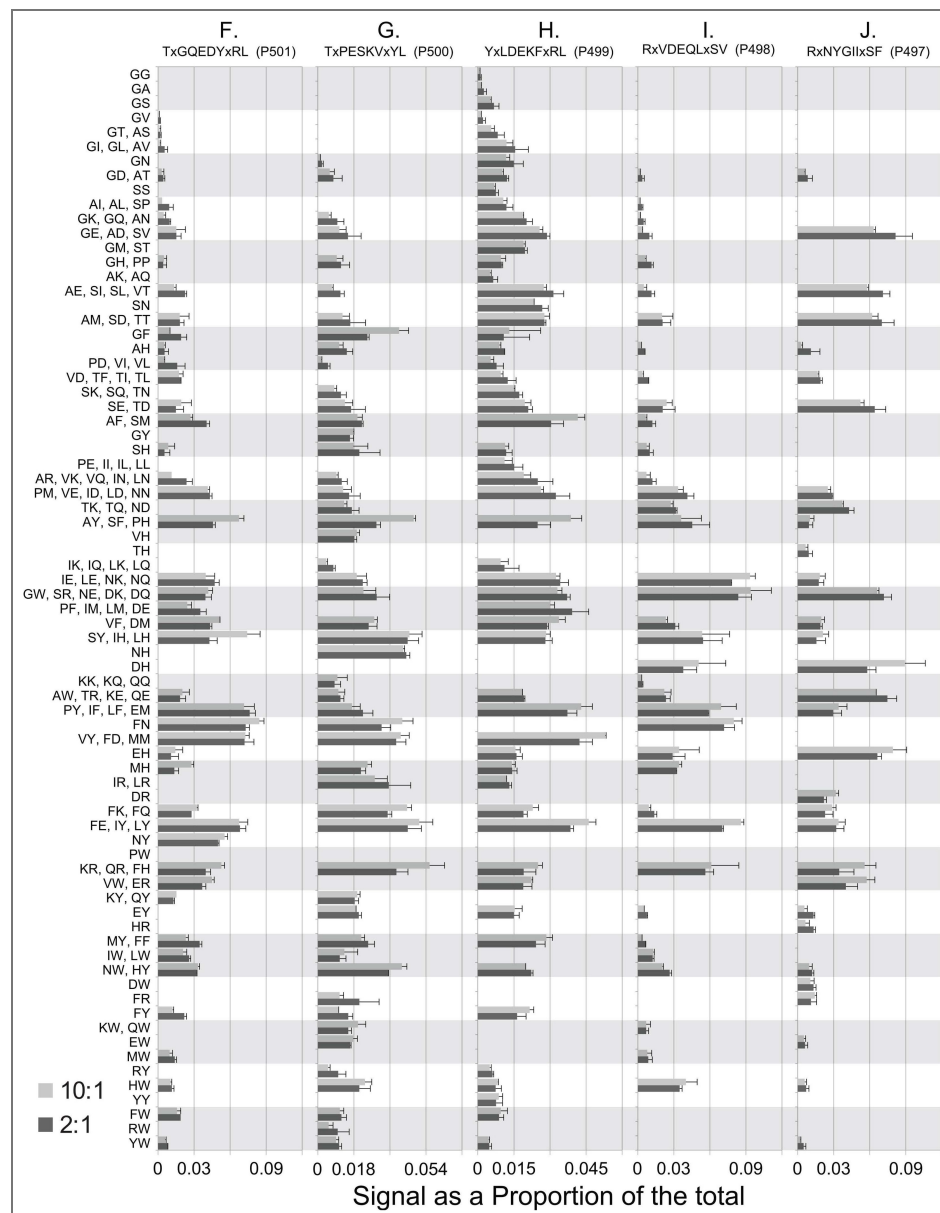


Figure 10.

Panels F-J (continued from Fig. 9). Assembly assays with double-substitution peptide libraries based on five 11mer peptides (GHPDWYYGETW and four derivatives of the 11mer GHAAEYGAETL) and five 10mers (all from the original peptides found by gas phase sequencing, including the original REVDEQLLSV) show different preferences for particular amino acid combinations at P_2 , P_3 , P_{C-3} and P_{C-2} positions binding to BF2*21:01, with more combinations found for 10mer peptides. Bacterial-expressed heavy chain and β_2 -microglobulin were mixed with peptide libraries (19 amino acids, all except Cys, at one position, indicated by x in the peptide sequence) at two molar ratios (1:2:10 as usual, and 1:2:2 to control for strong binding peptides) and separated by SEC as described in the legend to Fig. 2 and separated immediately by SEC; monomer peaks were collected, concentrated, treated with acid, lyophilised and analysed by MALDI-TOF. Bar charts show proportion of total signal (with error bars when present showing standard error around the mean for separate refolding experiments) on x-axis for mass/charge (m/z) positions corresponding to peptides with particular amino acid combinations on the y-axis (single letter code; Ile and Leu, Gln and Lys, and amino acids present in P_2 versus P_{C-2} positions were indistinguishable). Panels: F. TxGQEDYxRL, G. TxPESKVxYL, H. YxLDEKFxRL, I. RxVDEQLxSV, J. RxNYGIIxSF.

results by LC-MS/MS were amalgamated to give values that could be directly compared to the MALDI-TOF results (Figs. 11 [11](#)-13 [13](#)). Many combinations found by MALDI-TOF were also found by LC-MS/MS (ranging from 30-56.5% of total) with only a few found by MALDI-TOF but not by LC-MS/MS (ranging from 0-13% of total), while many more combinations were found by LC-MS/MS only (ranging from 30.4-68.3% in groups that would have been equivalent to the MALDI-TOF identifications). The proportions of each combination by MALDI-TOF were higher than with LC-MS/MS, due to having fewer combinations in total. Thus, LC-MS/MS was more accurate and more sensitive than MALDI-TOF.

The LC-MS/MS results broadly follow those from the MALDI-TOF analysis (Figs. 11 [11](#)-13 [13](#)). Overall, there were 159 combinations at P_2 and P_{c-2} found for double-substitution libraries based on four peptide backbones, with several (EE, HA, HD, HE, HF, HG, HH, HM, HN, HS and HY) found from all four libraries. However, there were more combinations found for the high affinity 10mer RxVDEQLxSV (75 combinations) than for the original 11mer GxAEEYGxTL (60), the derivative GxVEEYGxTL (44) and the low affinity 11mer GxPDWYYGxTW (23). The frequencies of amino acids in GxAEEYGxTL and RxVDEQLxSV (Fig. 14 [14](#)) show that both prefer His at P_2 (despite the original 10mer having Glu at P_2) with a smattering of other amino acids including Glu and Ser. However, there were differences at P_{c-2} , with the 11mer having Asp followed by Glu, His, Phe, Lys, Gln, Met and others at lower frequencies, but with the 10mer having nearly equal frequencies of Asp, Glu, Phe and Met, followed by Ile, Asn and others at lower frequencies. Moreover, the low affinity 11mer GxPDWYYGxTW has almost only His at P_2 , but a variety of amino acids at P_{c-2} , most of which are poorly represented at P_{c-2} in GxAEEYGxTL and RxVDEQLxSV (Fig. 15 [15](#)). Thus, the backbone of the peptide has effects on the preferences at P_{c-2} .

The differences between GxAEEYGxTL and GxVEEYGxTL for LC-MS/MS and MALDI-TOF support the idea that P_3 is an important position within the peptide (Figs. 8 [8](#)-13 [13](#)) despite not being an anchor residue (Chappell et al., 2015 [15](#); Koch et al., 2008). In particular, the predominant His at P_2 in the original 11mer with Ala at P_3 changes to a more equal amount of His and Glu with significant Ala at P_2 in the derivative with a Val at P_3 . Both LC-MS/MS and MALDI-TOF showed many combinations for P_3 and P_{c-3} in the original 11mer GHxEEYGxETL and the derivative GExEEYGxLTL; the frequencies of amino acids at P_3 were similar (with an increase Gly in the derivative), but several differences were found for P_{c-3} (including decreases in Lys, Phe, Arg and Ser in the derivative) (Fig. 16 [16](#)).

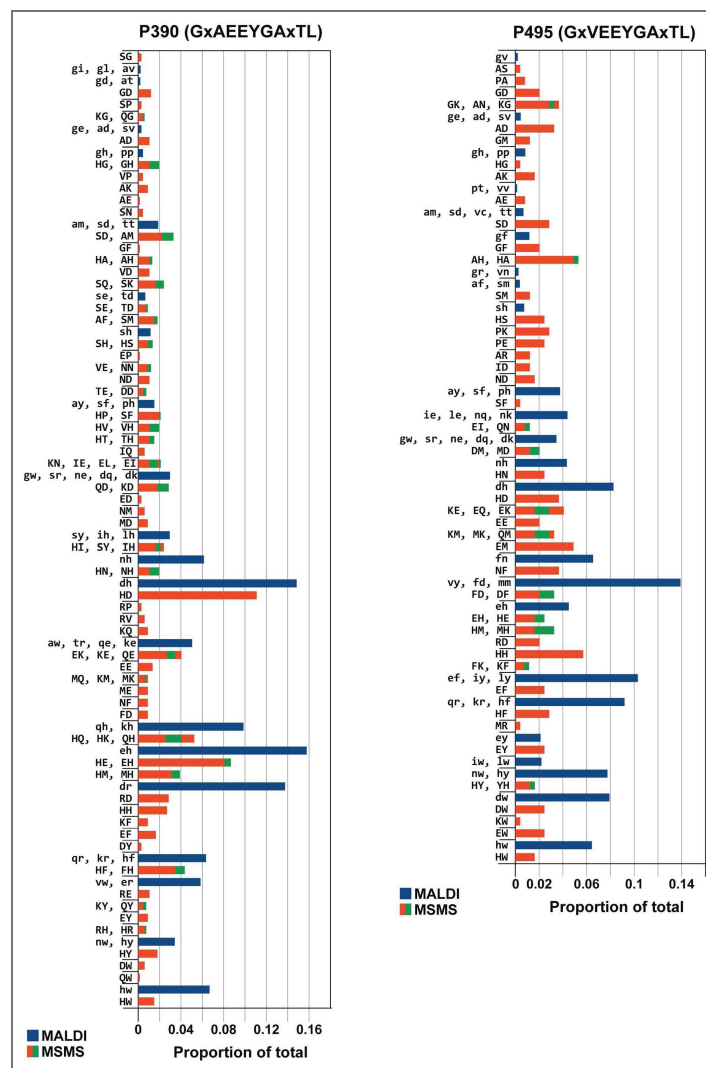


Figure 11.

Panels A-B. Comparison of assembly assays of double-substitution libraries by analysis with MALDI-TOF and with LC-MS/MS give similar results, with different preferences for particular amino acid combinations at P_2 , P_3 , P_{C-3} and P_{C-2} positions binding to BF2*21:01, with more combinations for 10mer peptides. Bacterially-expressed heavy chain and β_2 -microglobulin were mixed with peptide libraries (19 amino acids, all except Cys, at two positions indicated by x in the peptide sequence) at two molar ratios (1:2:10 as usual, and 1:2:2 to control for strong binding peptides) and separated by SEC as described in the legend to Fig. 2; monomer peaks were collected, concentrated, treated with acid, lyophilised and analysed. Separate experiments were analysed by MALDI-TOF (blue bars, amino acids in single letter code with small letters, order not meaningful) and by LC-MS/MS (orange and green stacked bars, amino acids in single letter code with capital letters, first letter corresponding to P_2 or P_3 and second letter corresponding to P_{C-2} or P_{C-3} , depending on peptide). Bar charts show proportion of total signal (with error bars when present showing standard error around the mean for separate refolding experiments) on x-axis for mass/charge (m/z) positions corresponding to peptides with particular amino acid combinations on the y-axis. Panels: A. GxAEEYGxTL, B. GxVEEYGxTL.

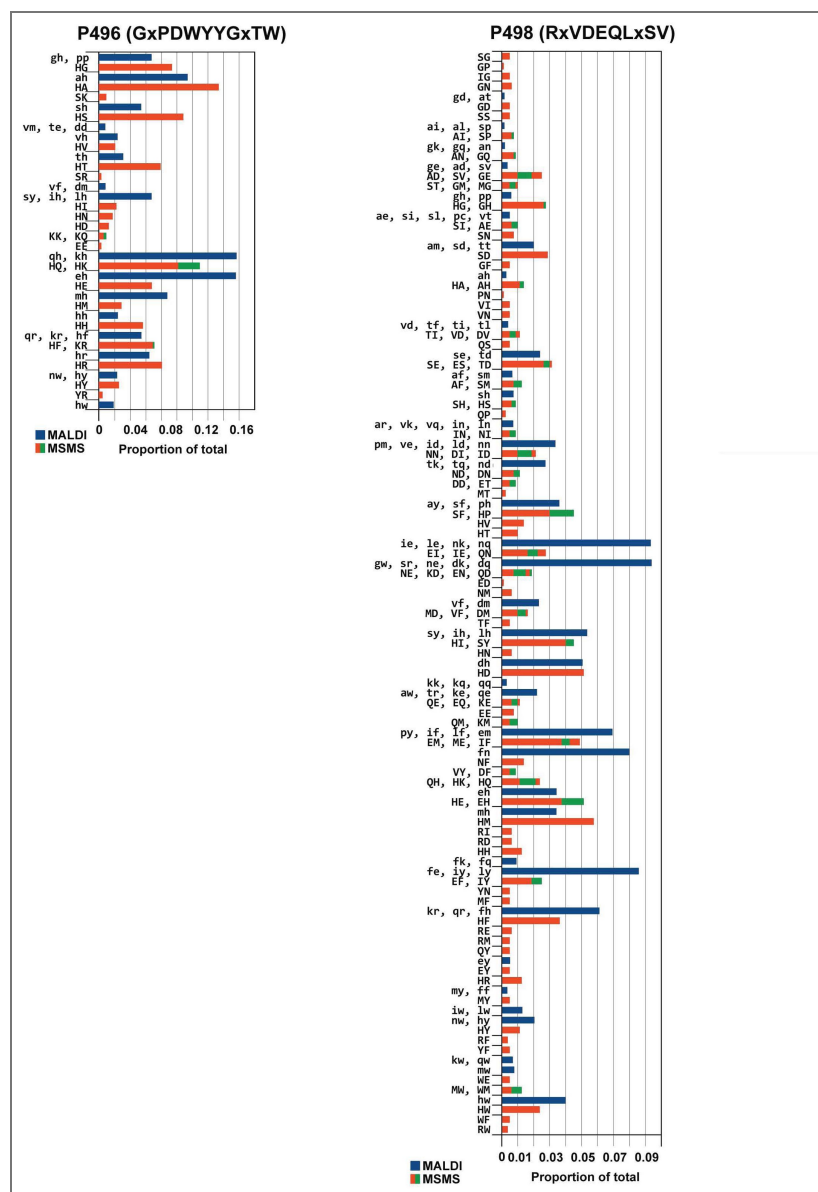


Figure 12.

Panels C-D (continued from Fig. 11). Comparison of assembly assays of double-substitution libraries by analysis with MALDI-TOF and with LC-MS/MS give similar results, with different preferences for particular amino acid combinations at P_2 , P_3 , P_{C-3} and P_{C-2} positions binding to BF2*21:01, with more combinations for 10mer peptides. Bacterially-expressed heavy chain and β_2 -microglobulin were mixed with peptide libraries (19 amino acids, all except Cys, at two positions indicated by x in the peptide sequence) at two molar ratios (1:2:10 as usual, and 1:2:2 to control for strong binding peptides) and separated by SEC as described in the legend to Fig. 2; monomer peaks were collected, concentrated, treated with acid, lyophilised and analysed. Separate experiments were analysed by MALDI-TOF (blue bars, amino acids in single letter code with small letters, order not meaningful) and by LC-MS/MS (orange and green stacked bars, amino acids in single letter code with capital letters, first letter corresponding to P_2 or P_3 and second letter corresponding to P_{C-2} or P_{C-3} , depending on peptide). Bar charts show proportion of total signal (with error bars when present showing standard error around the mean for separate refolding experiments) on x-axis for mass/charge (m/z) positions corresponding to peptides with particular amino acid combinations on the y-axis. Panels: C. GxPDWYYGxTW, D. RxVDEQLxSV.

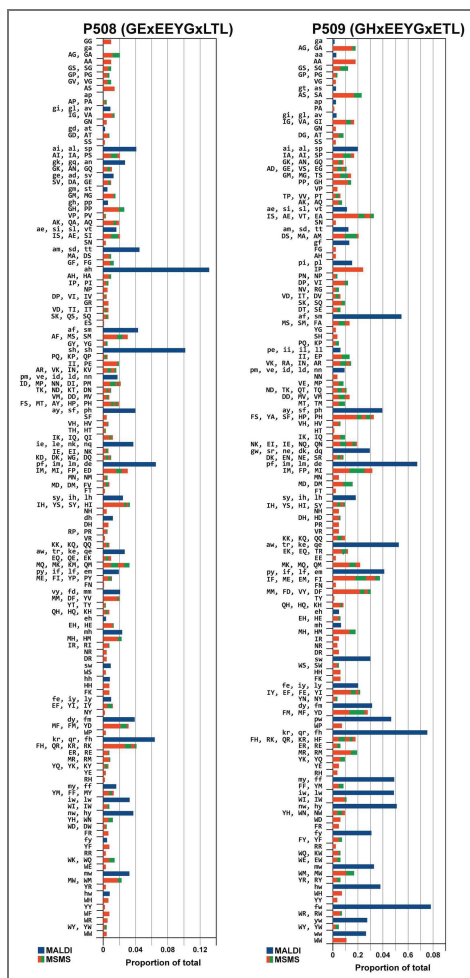


Figure 13.

Panels E-F (continued from Figs. 11 and 12). Comparison of assembly assays of double-substitution libraries by analysis with MALDI-TOF and with LC-MS/MS give similar results, with different preferences for particular amino acid combinations at P_2 , P_3 , P_{C-3} and P_{C-2} positions binding to BF2*21:01, with more combinations for 10mer peptides. Bacterially-expressed heavy chain and β_2 -microglobulin were mixed with peptide libraries (19 amino acids, all except Cys, at two positions indicated by x in the peptide sequence) at two molar ratios (1:2:10 as usual, and 1:2:2 to control for strong binding peptides) and separated by SEC as described in the legend to Fig. 2; monomer peaks were collected, concentrated, treated with acid, lyophilised and analysed. Separate experiments were analysed by MALDI-TOF (blue bars, amino acids in single letter code with small letters, order not meaningful) and by LC-MS/MS (orange and green stacked bars, amino acids in single letter code with capital letters, first letter corresponding to P_2 or P_3 and second letter corresponding to P_{C-2} or P_{C-3} , depending on peptide). Bar charts show proportion of total signal (with error bars when present showing standard error around the mean for separate refolding experiments) on x-axis for mass/charge (m/z) positions corresponding to peptides with particular amino acid combinations on the y-axis. Panels: E. GExEEYGxLTL, F. GHxEEYGxETL.

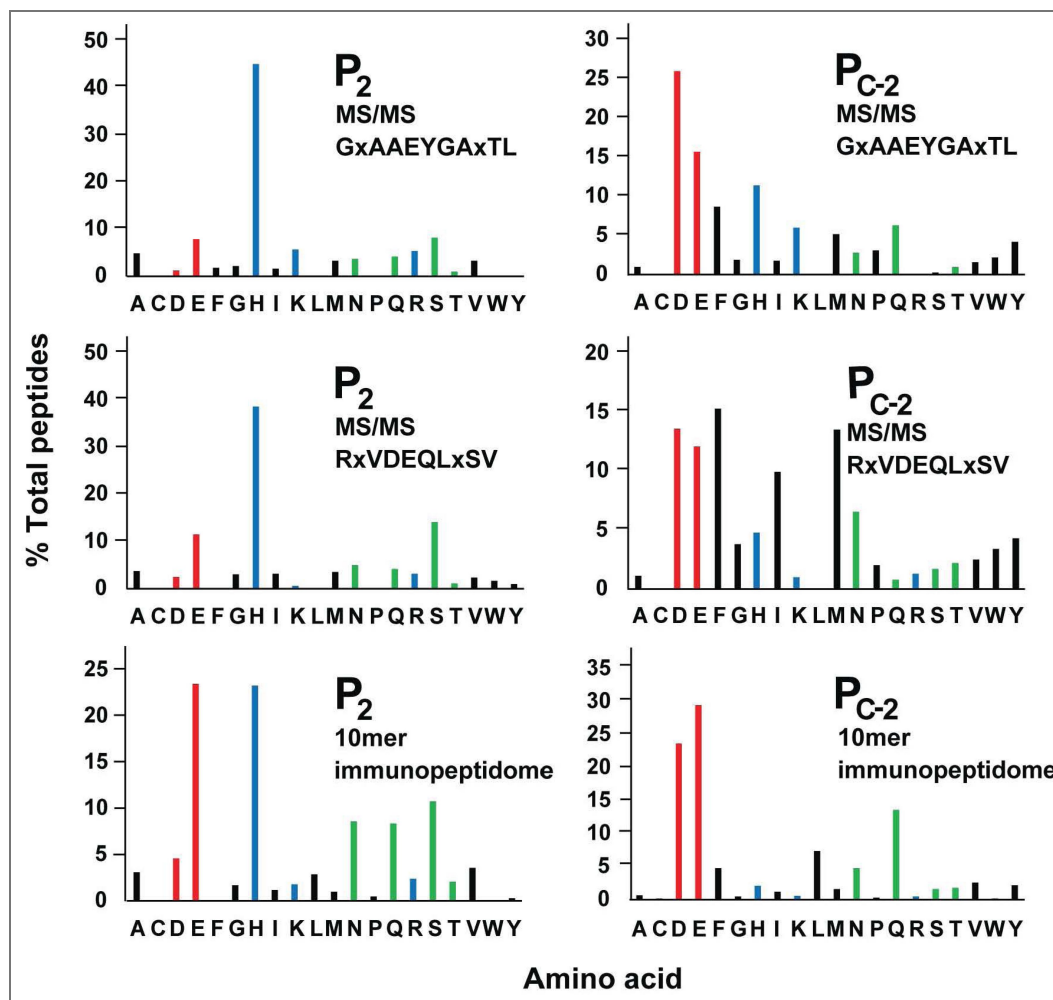


Figure 14. The frequencies of particular amino acids found at positions P_2 and P_{C-2} of peptides bound to class I molecules from the B21 haplotype vary considerably, as assessed by LC-MS/MS of peptides from two double-substitution libraries with BF2*21:01 and from immunopeptidomics of a B21 cell line.

For assembly assays, bacterial-expressed heavy chain and β_2 -microglobulin were refolded in vitro with either GxAAEYGAxTL or RxVDEQLxSV (single letter code; x means roughly equal proportions of 19 naturally-occurring amino acids, all except Cys), and separated by SEC; monomer peaks were collected, concentrated, treated with acid, and lyophilised. For immunopeptidomics, class I molecules from detergent-solubilised AVOL-1 cells were isolated by affinity chromatography with monoclonal antibody F21-2. Samples were analysed by LC-MS/MS, with bar graphs showing the percentage of sequences with different amino acids (single letter code) at P_2 or P_{C-2} found by mass spectroscopy; immunopeptidomic data for 10mer peptides only.

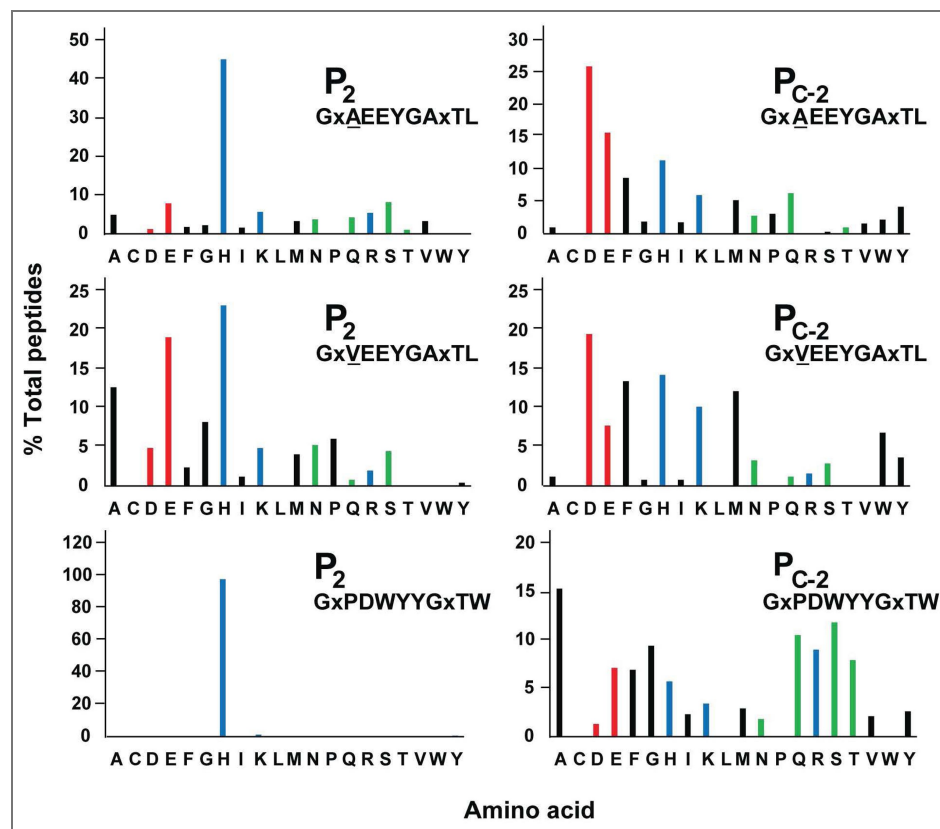


Figure 15. The frequencies of particular amino acids found at positions P_2 and P_{C-2} in peptides bound to BF2*21:01 vary considerably and are affected by the amino acid present at P_3 , as assessed by LC-MS/MS of peptides from assembly assays with three double-substitution libraries.

Bacterially-expressed heavy chain and β_2 -microglobulin were refolded in vitro with either $GxAEEYGxTL$, $GxVEEYGxTL$ or $GxPDWYYGxTW$ (single letter code; x means roughly equal proportions of 19 naturally-occurring amino acids, all except Cys) and separated by SEC as described in the legend to Fig. 2; monomer peaks were collected, concentrated, treated with acid, lyophilised and analysed by LC-MS/MS, with bar graphs showing the percentage of sequences with different amino acids (single letter code) at P_2 or P_{C-2} found by mass spectroscopy.

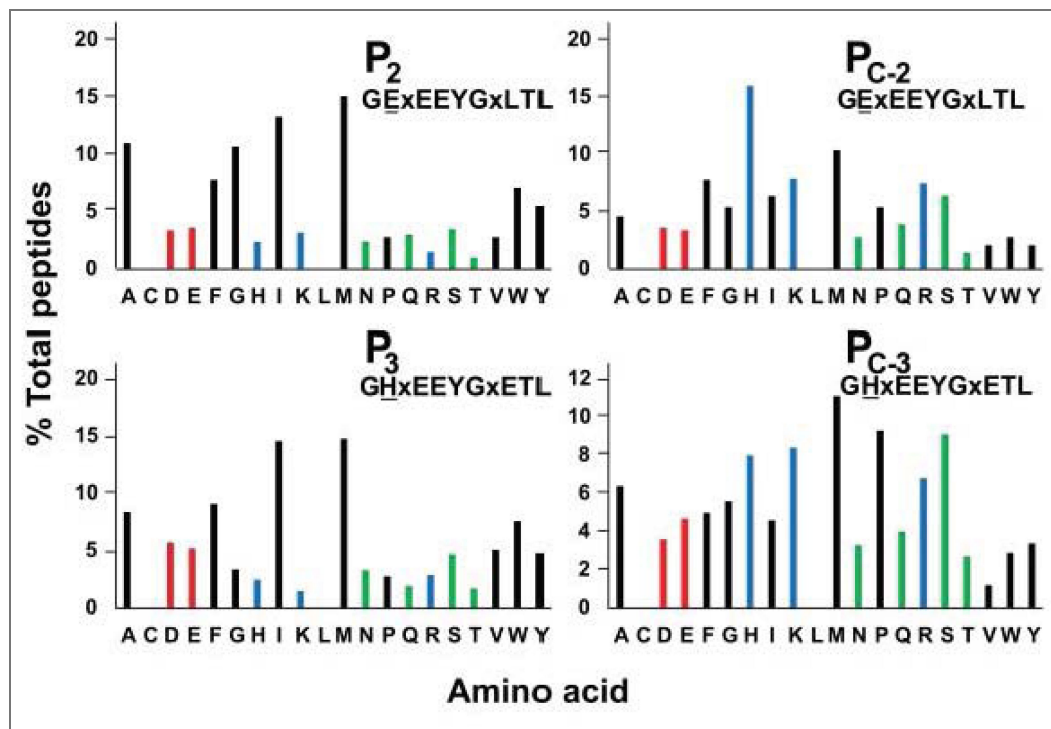


Figure 16. The frequency of particular amino acids found at positions P_3 and P_{C-3} in peptides bound to BF2*21:01 is affected by the amino acids present at P_2 and P_{C-2} , as assessed by LC-MS/MS of peptides from assembly assays with two double-substitution libraries.

Bacterially-expressed heavy chain and β_2 -microglobulin were refolded in vitro with either GExEEYGxLTL or GHxEEYGxETL (single letter code; x means roughly equal proportions of 19 naturally-occurring amino acids, all except Cys) and separated by SEC as described in the legend to Fig. 2; monomer peaks were collected, concentrated, treated with acid, lyophilised and analysed by LC-MS/MS, with bar graphs showing the percentage of sequences with different amino acids (single letter code) at P_3 or P_{C-3} found by mass spectroscopy.

Immunopectidomics shows wide variation but only a few predominant combinations in anchor residues P₂ and P_{C-2}

In comparison with in vitro assembly assays, the peptides loaded within a cell are subject to quality control mechanisms, so the patterns found from in vitro and cellular analyses need not be the same. Immunopectidomic analysis (Purcell et al., 2019) was used for the peptides eluted from class I molecules expressed in chicken cell lines, with a wide variety of peptides found for B21 compared to B19 (Chappell et al., 2015). There are two class I molecules in these preparations, but the expression of BF1 is much lower than BF2 (Shaw et al., 2008; Wallny et al., 2006). Immunopectidomics of the B21 cells identified 3818 peptides with a range from 5mers to 50mers (Fig. 17 – Source Data 1), but the most frequent were found as 8, 9, 10, 11 and 12mers (2972 peptides, 77.8% of total). For each of these lengths, the peptide motifs determined by Gibbs clustering were complex (Fig. 17), as was found by in vitro assays. The 10mers were by far the most frequent (1470 peptides, 49.5% of 8-12mers) (Fig. 17), consonant with the wider variety of anchor residues and the greater thermostability noted above (Figs. 6-16). Peptides shorter than 8mers are conceivably proteolytic products, but at least some of the peptides longer than 12mers may have been bound to class I molecules, either by extra-long bulging in the middle of the peptide (Burrows et al., 2006; Liu et al., 2012) or by extending out of the groove at the C-terminus as has been found for other chicken class I molecules (Xiao et al., 2018).

By immunopectidomics, the frequencies of amino acids at key positions were similar but not identical for different peptide lengths (Fig. 18). For P₂, the 10mers had roughly equal amounts of Glu and His followed by roughly equal levels of Asn, Gln and Ser and with other amino acids at lower levels. The 9mers and 11mers were enriched for His at P₂ compared to 10mers, consistent with lower affinities of 9mers and 11mers not able to tolerate other amino acids so well. For P_{C-2}, peptides of all three lengths prefer Asp and Glu, although they are roughly equal for 10mers, Asp favoured for 9mers and Glu favoured for 11mers; some other amino acids are more frequent in 10mers than 9mers or 11mers. The patterns for 12mers look like 9mers and 11mers, but for 8mers look more like 10mers (Fig. 18); studies in other haplotypes suggest that these 8mers may be at least in part from the BF1 molecule (data not shown). However, peptides of all five lengths overwhelmingly have Leu at P_C (Fig. 18).

The analyses to resolve whether the peptides greater than 12mers bulged in the middle or hung out the C-terminal end gave equivocal results (Fig. 19). The 204 peptides from 13-15 amino acids in length had more His at P₂ and Leu at P_C than other amino acids, but unlike the 12mers, the proportion of His at P₂ was only marginally above many of the other amino acids and there was almost as much Lys as Leu at P_C, with many other amino acids at significant amounts. The high levels of Asp, Glu and Leu at P_{C-2} found in 12mers might be expected to be located at P_{C-2} regardless of length if the peptides bulge in the middle, but might be expected at P₈ regardless of length if a 10mer length is preferred and the rest of the peptide hangs out the end. Asp, Glu and Leu were high at both P_{C-2} and at P₈, but many other amino acids were found there as well. It is possible that the several modes of binding contribute to the overall result, which were not possible to disentangle by these analyses.

Overall, the 442 peptides with lengths of 16 amino acids and above were even more different from the 12mers, with Lys favoured at both P₂ and at P_C but with several other amino acids as well, and with the proportions of Asn, Glu and Ser at P_C increasing with increasing length.

The 10mer peptides had three groups of amino acids at P₂ in roughly equal amounts: acidic (predominantly Glu), basic (predominantly His) or polar (roughly equally Asn, Gln and Ser with some Thr), with the remainder having hydrophobic amino acids at P₂ (Fig. 20). The 10mers with acidic amino acids at P₂ favoured Gln at P_{C-2} followed by Leu with small amounts of other amino acids, while the 10mers with basic, polar and hydrophobic amino acids at P₂ favoured Asp and Glu at P_{C-2} with low levels of other amino acids, consonant with the co-variation at positions P₂ and P_{C-2}.

2.

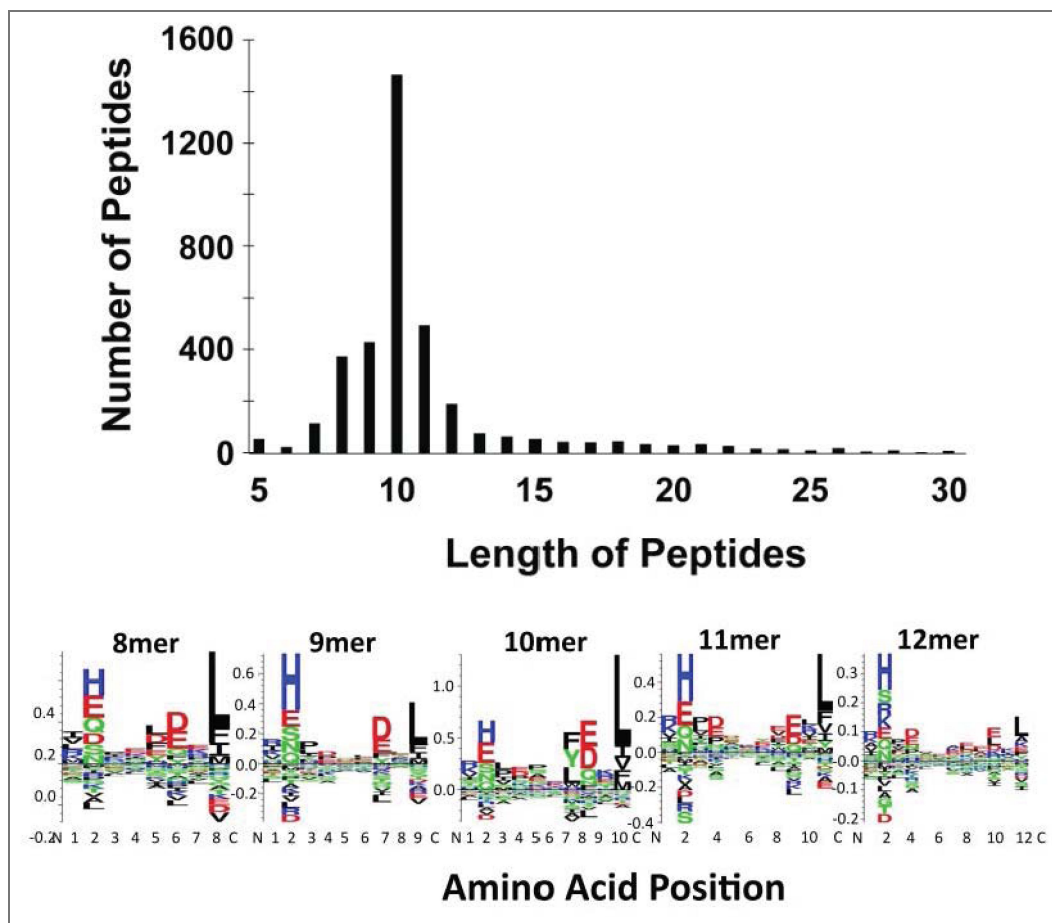


Figure 17. Most peptides found by immunopeptidomics of a B21 cell line were between 8 and 12 amino acids in length, with the most frequent being 10mers, all of which had complex peptide motifs.

For immunopeptidomics, class I molecules from detergent-solubilised AVOL-1 cells were isolated by affinity chromatography with monoclonal antibody F21-2. Samples were analysed by LC-MS/MS, with bar graphs showing the number of sequences with different lengths as found by mass spectroscopy. Peptide motifs were determined by Gibbs clustering.

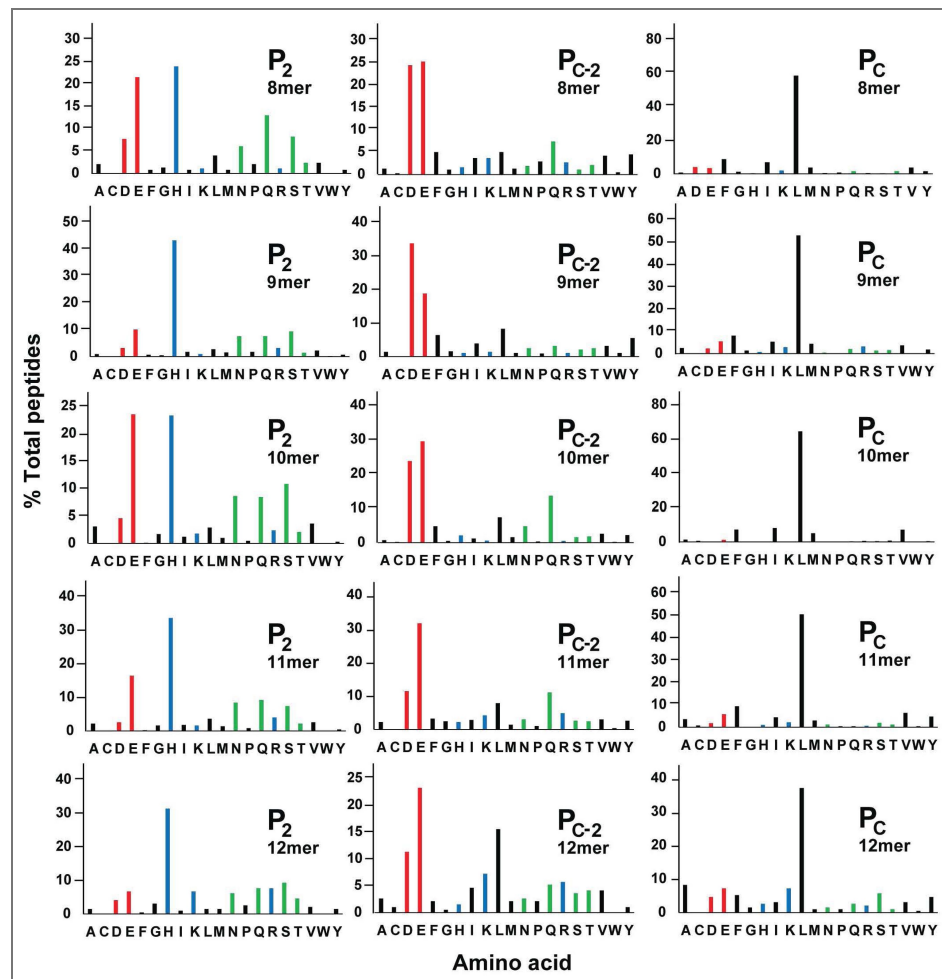


Figure 18. The frequency of particular amino acids found at positions P_2 , P_{C-2} and P_C in peptides bound to class I molecules from a B21 cell line varies depending on length of peptide, with Leu predominant at P_C for all lengths, and with 8mers more like 10mers at P_2 and P_{C-2} .

Peptides identified by immunopeptidomics as in Fig. 17, with bar graphs showing the percentage of sequences with different amino acids (single letter code) at P_2 , P_{C-2} and P_C found by mass spectroscopy.

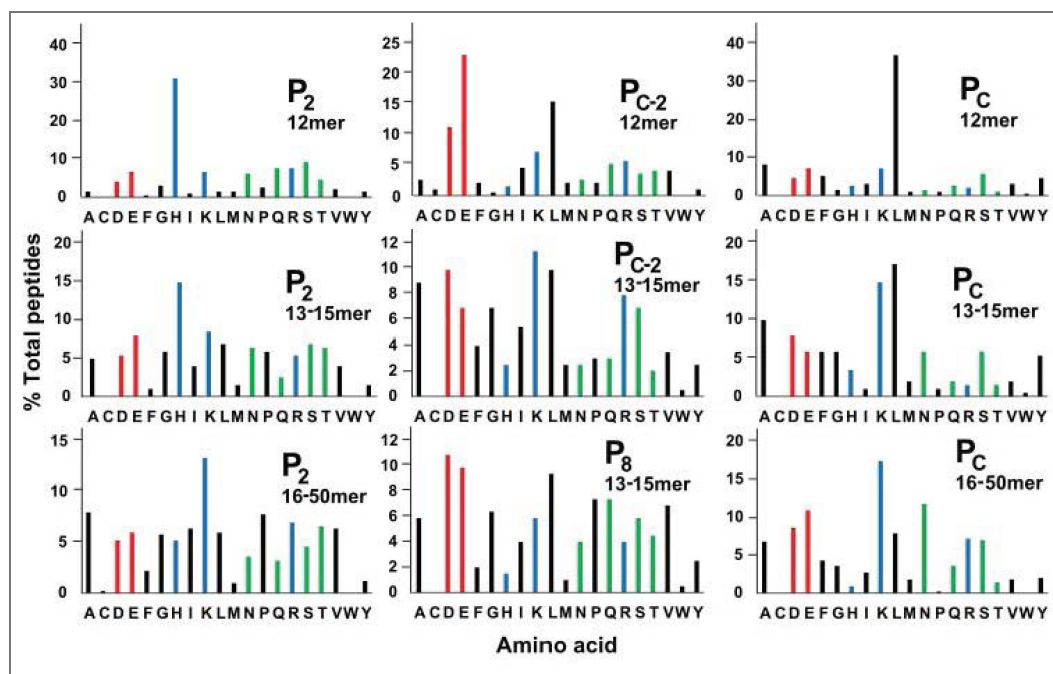


Figure 19. The frequency of particular amino acids found at positions P_2 , P_{C-2} and P_C in peptides bound to class I molecules from a B21 cell line varies depending on length of peptide, but patterns found in peptides up to 12mers are not so obvious in longer peptides.

Peptides identified by immunopeptidomics as in Fig. 17, with bar graphs showing the percentage of sequences with different amino acids (single letter code) at P_2 , P_{C-2} and P_C found by mass spectroscopy.

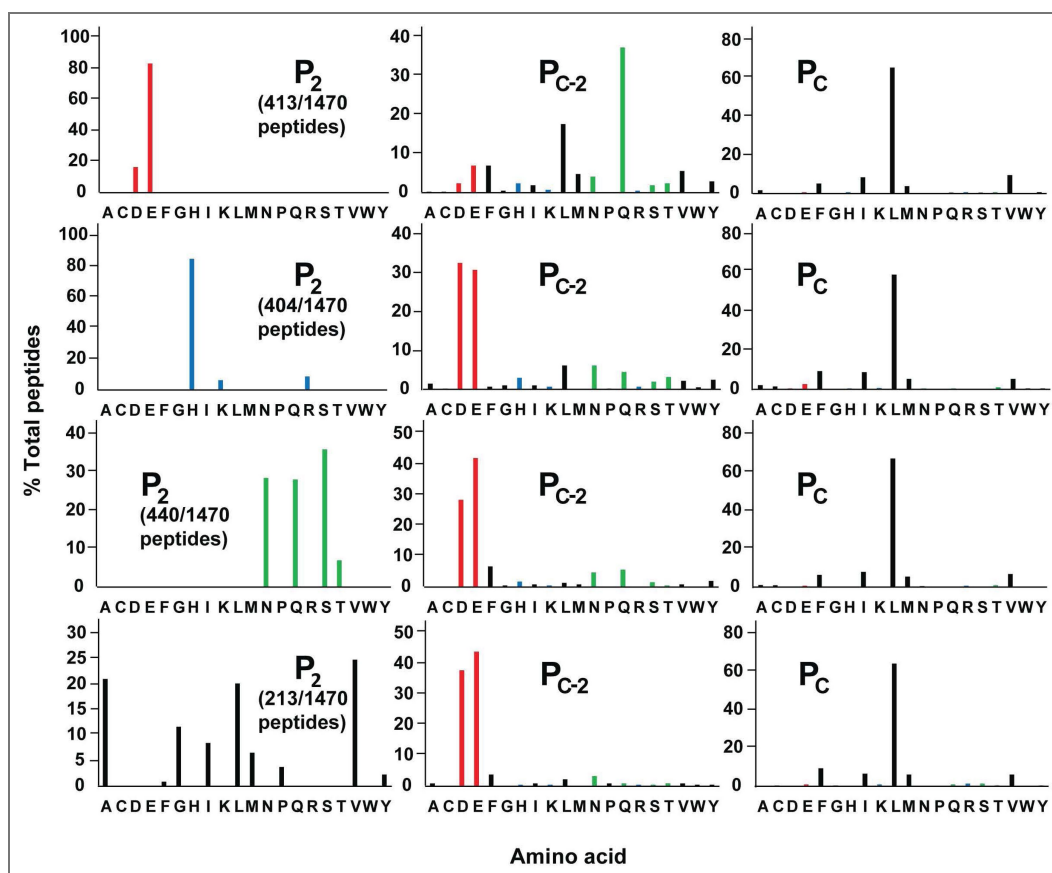


Figure 20. The frequencies of particular amino acids found at position P_{C-2} in peptides bound to class I molecules from a B21 cell line vary depending on whether the amino acids at P_2 are acidic (Asp and Glu), basic (His, Lys and Arg), polar (Asn, Gln, Ser and Thr) or hydrophobic (Ala, Cys, Phe, Gly, Ile, Leu, Met, Pro, Val, Trp and Tyr), but Leu is found overwhelmingly at P_C in all peptides.

Peptides identified by immunopeptidomics as in Fig. 17, with bar graphs showing the percentage of sequences with different amino acids (single letter code) at P_2 , P_{C-2} or P_C found by mass spectroscopy.

There were differences between the amino acid frequencies by immunopeptidomics and by assembly assays. In contrast to the predominant His at P₂ for the assembly assays of both the 11mer GxAEEYGAXTL and 10mer RxVDEQLxSV, immunopeptidomics showed roughly equal Glu and His along with lesser amounts of Asn, Gln and Ser, followed by other amino acids (Fig. 14). In comparison to the complex patterns at P_{C-2} for the assembly assays, immunopeptidomics had predominant Asp and Glu with lesser amounts of Gln, Leu, Asn and Phe, followed by other amino acids (Fig. 14).

Co-variation between amino acids at P₂ and P_{C-2} might be expected to reduce the number of combinations that bind to BF2*21:01, but 238 combinations were found for 9-11mers (273 for 8-12mers) out of 400 possible (Fig. 21). However, most of these combinations were found at very low levels, with the number of combinations found at greater than 1% frequency being only 19 for 9mers, 27 for 10mers and 17 for 11mers (Fig. 22); again 8mers were more like 10mers, perhaps due to the presence of peptides from BF1*2102. Of 36 high frequency combinations, 18 were shared among at least two of the 9mers, 10mers and 11mers (29 of 48 combinations if 8mers and 12mers are included). The combination His at P₂ and Glu at P_{C-2} was predominant among 11mers at greater than 12% of total peptides, but His and Asp at nearly 16% was even more than His and Glu at 12% for the 9mers. A more even distribution was found for 10mers, with Glu and Gln at roughly 8%, His and Glu at over 7%, and His and Asp at nearly 7%; again 8mers were more like 10mers. Overall, those combinations above 1% frequency corresponded to 60% of 9mers, 71% of 10mers and 57% of 11mers, with the top two 9mers amounting to 27%, the top three 10mers to 22%, and the top 11mer to 12.4%. Thus, while a majority of combinations at P₂ and P_{C-2} is possible, only a few were found frequently.

Immunopeptidomics shows wide variation and complex patterns for P₃ and P_{C-3} that are affected by P₂ and P_{C-2}

Analysis from the immunopeptidomic data supports the interpretation from the assembly assays that the amino acids allowed at the anchor residues (in the sense of binding into deep pockets of the MHC molecule) at P₂ and P_{C-2} may be affected by amino acid sidechains from positions that are not anchor residues. In particular, the amino acids at P₃ and P_{C-3}, which do not contact the MHC molecule at all (Chappell et al., 2015; Koch et al., 2008), are found at wildly varying proportions depending on peptide length (Fig. 23). Pro at P₃ was found at high frequency (24% for 9mers, 17% for 10mers, 18% for 11mers) as well as Leu (10% for 9mers, 19% for 10mers, 13% for 11mers), with significant amounts of Ala, Ile, Met, Ser and Val for 10mers (and less for 9mers). At P_{C-3}, Phe, Leu and Tyr were frequent (each at about 20%) with a smattering of other residues in 10mers, while in addition Ala, Asp, Glu, Ile, Lys, Pro and Val were found half as often in the 9mer, and Glu was twice as likely as Ala, Asp, Phe, His, Ile, Leu, Ser, Val and Trp in the 11mers. The same amino acids were found for 8mers but at different proportions (Fig. 23), which may indicate binding to BF1 rather than BF2.

A more detailed look at the immunopeptidomic data suggests that the amino acids allowed at positions P₃ and P_{C-3} may depend on the amino acids at P₂ and P_{C-2} (Figs. 24, 25). For example, acidic residues (Asp and Glu) at P₃ are found at high levels only in 11mers with Tyr at P₂, in 10mers with Phe (and less so for Tyr) at P₂, and in 9mers with several amino acids at P₂ (Fig. 24). In contrast, basic residues at P₃ are nearly all found in 11mers with Phe at P₂, in over 80% of 9mers with Ile, Leu and Tyr at P₂, but much more evenly distributed in 10mers. Large hydrophobic residues (Phe, Ile, Leu, Met and Tyr) at P₃ are found at high levels in many peptides, but not 9mers with Ala, Phe, Gly, Ile, Lys and Met at P₂, in 10mers with Phe and Pro at P₂, and in 11mers with Ala, Phe and Tyr at P₂ (as well as Cys and Trp, but there are not many peptides total for these residues). Similarly, there are few if any peptides with small hydrophobic residues (Ala, Cys, Gly and Ser) at P₃ for 9mers with Asp, Glu and Tyr, and for 11mers with Phe, Lys, Pro and Tyr. The kinked amino acid Pro at P₂ is found in nearly 70% of 10mers with Pro at P₃, and strong (but not so extreme) skewing is found for 9mers and 11mers. For example, for 9mers with Pro at P₃ there are high levels of Phe, Gly, His, Met, Asn, Ser and Thr at P₂, but none with Lys, Leu, Arg or Tyr. Such unequal distributions of P_{C-3} dependent on P_{C-2} are found for 9mers, 10mer and 11mers as

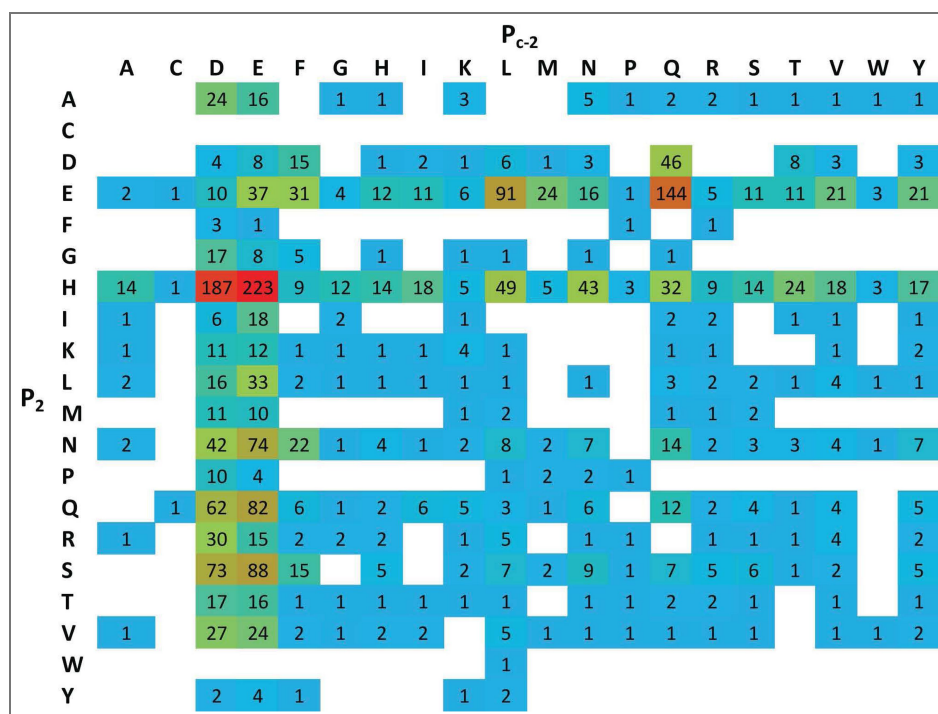


Figure 21. Many different combinations of amino acids at positions P₂ and P_{C-2} in peptides bound to class I molecules from a B21 cell line are found by immuno-peptidomics, but only a few are found frequently.

Peptides (9, 10 and 11mers) identified by immuno-peptidomics as in Fig. 17, with heat map showing combinations of amino acids (single letter code, with P₂ on left and P_{C-2} on top) at intersections, with empty cells showing no peptides found, and colours showing numbers of peptides found highlighted with low frequency in blue to high frequency in red.

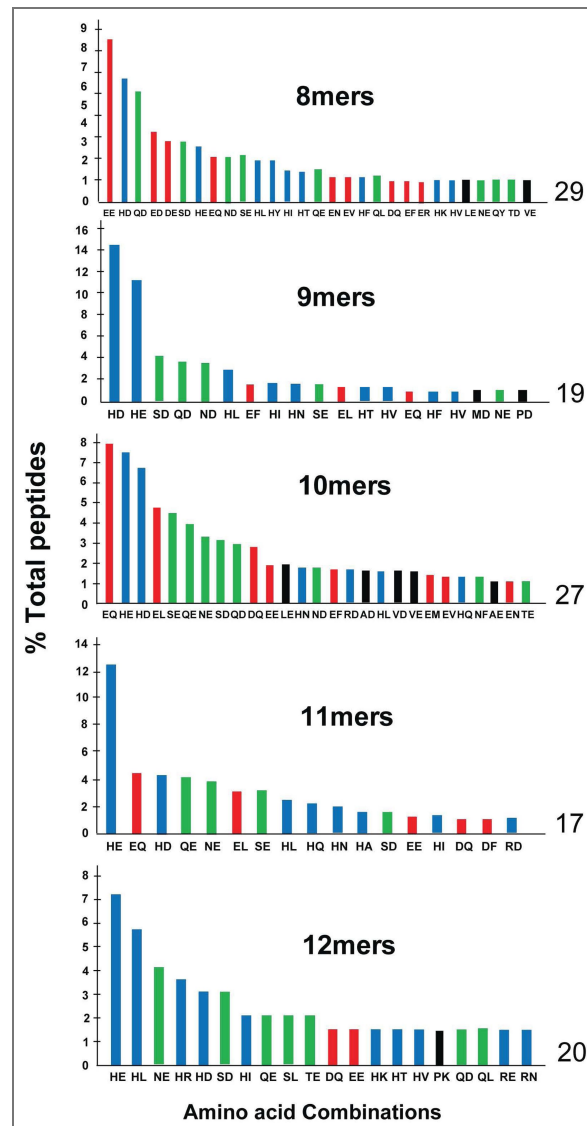


Figure 22. Only a few combinations of amino acids at positions P_2 and P_{C-2} in peptides bound to class I molecules from a B21 cell line are found at frequencies of 1% or higher.

Peptides identified by immunopeptidomics as in Fig. 17, with bar graphs showing the percentage of sequences with different combinations of amino acids at P_2 and P_{C-2} (single letter code, with P_2 first and then P_{C-2}) found by mass spectroscopy.

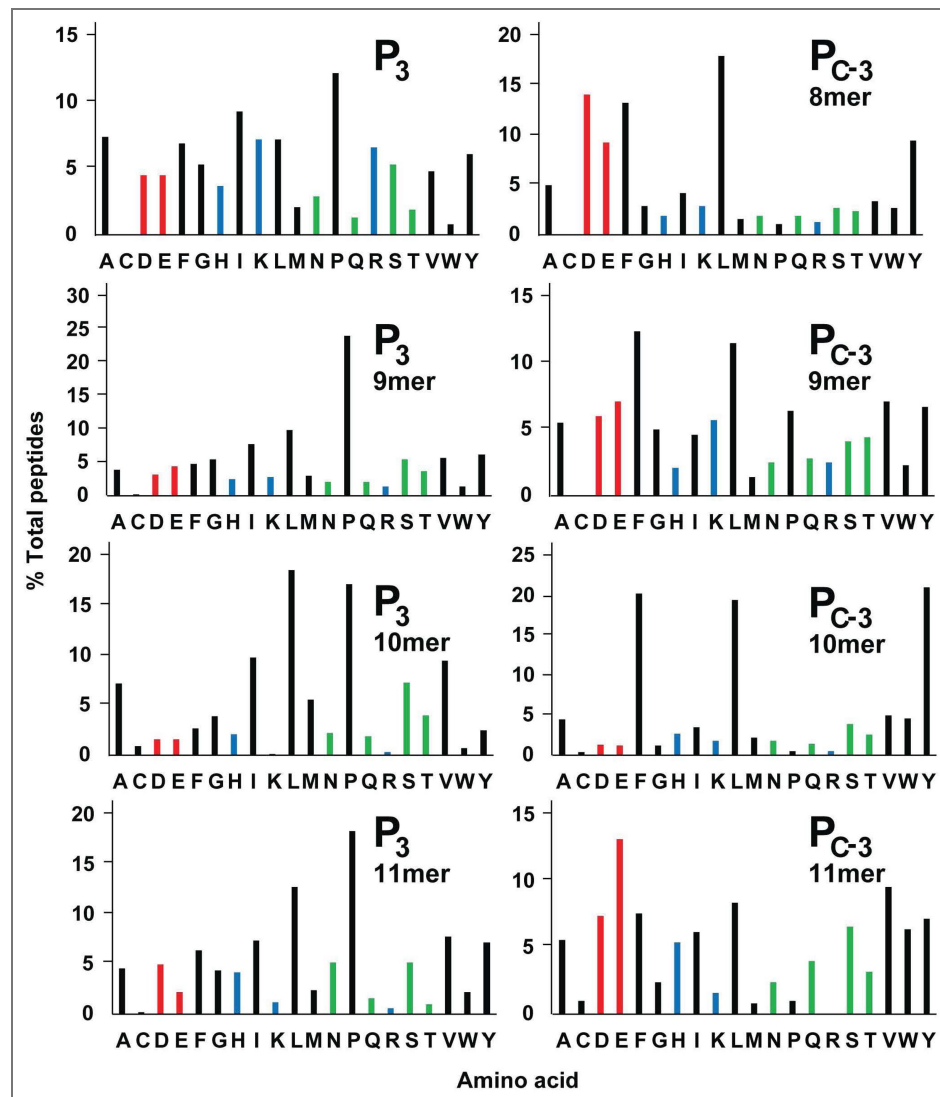


Figure 23. The frequencies of particular amino acids found at positions P_3 and P_{C-3} in peptides bound to class I molecules from a B21 cell line vary depending on length of peptide.

Peptides identified by immunopeptidomics as in Fig. 17, with bar graphs showing the percentage of sequences with different amino acids (single letter code) at P_3 or P_{C-3} found by mass spectroscopy.

well (Fig. 25 [↗](#)). Particularly striking are the extreme preferences in 9mers for Arg at P_{C-2} with acidic residues at P_{C-3} , Pro at P_{C-2} with basic residues at P_{C-3} , and Ala at P_{C-2} with Pro at P_{C-3} . Thus, whether a particular peptide binds BF2*21:01 depends not only on the anchor residues at P_2 , P_{C-2} and P_C , but also on residues from the rest of the peptide, such as P_3 and P_{C-3} .

Immuno-peptidomics shows a strong preference for Leu at P_C but co-variation with P_{C-2}

Assembly assays with GHAEYGAETx and REVDEQLLSx (Fig. 7 [↗](#)) show high and roughly equal amounts of Phe, Ile/Leu, Met and Val at P_C , but immuno-peptidomics shows Leu at P_C to be by far the most frequent, 50-60% frequency in 8-11mers (Fig. 18 [↗](#)). At much lower levels are found Ile, Phe, Met and Val, but in fact every single other amino acid is found at P_C by immuno-peptidomics. The reason underlying this strong preference for Leu at P_C in a living cell compared to assembly assays is not clear, but one possibility is co-variation with other peptide positions (unlike the assembly assays with fixed sequences).

Peptides with different amino acids at P_{C-2} can have much different preferences for amino acids at P_C (Figs. 26 [↗](#), 27 [↗](#)). The levels of Leu at P_C are greater than 50% for five amino acids in 9mers, 12 for 10mers and six for 11mers, with Asp, Glu and Gln at P_{C-2} having high Leu frequencies at P_C for all three lengths. In contrast, the percentages of Leu at P_C are low for Arg at P_{C-2} in 9mers, Arg and Lys for 10mers, and Arg, Lys and Ile for 11mers (Fig. 26 [↗](#)).

For many amino acids at P_{C-2} , there is a preponderance of Leu at P_C followed by other hydrophobic amino acids at P_C . For 10mers, the presence of Leu and other hydrophobic amino acids is almost complete. In contrast, 9mers with Ala, Phe, Leu, Met, Pro, Ser, Trp and Tyr as well as 11mers with Gly, His, Ile, Lys and Pro at P_C have greater than 35% acidic (Asp or Glu), basic (Arg, His or Lys) or polar (Asn, Gln, Ser or Thr) amino acids at P_{C-2} (Fig. 27 [↗](#)). For example, out of the 28 9mer sequences with Phe at P_{C-2} , only eight have Leu at P_C , but 13 (46.4%) have acidic, basic or polar amino acids. This compares to 70 10mer sequences with Phe at P_{C-2} , of which 41 have Leu at P_C and only five are acidic, basic or polar. Similarly, out of the 24 9mer sequences with Tyr at P_{C-2} , only five have Leu at P_C , but 15 (62.5%) have acidic, basic or polar amino acids; of 32 such sequences that are 10mers, 17 have Leu at P_C and only seven are acidic, basic or polar. Although there are only four 9mers, four 10mers and five 11mers with Pro at P_{C-2} , only one sequence each has Leu at P_C , while 50-75% are acidic, basic or polar. For sequences with Leu at P_{C-2} , 14 out of 36 9mers (39%), 12 out of 39 11mers (36%), and even 13 out of 109 10mers (12%) have acidic, basic or polar amino acids at P_C .

Such associations suggest that co-variation also occurs between the amino acids at P_{C-2} and P_C , presumably through the interactions with BF2*21:01 rather than directly. It is also clear (Fig. 24 [↗](#)) that the high frequency of 10mers and of Asp and Glu at P_{C-2} among peptides found on BF2*21:01 molecules favours overall the presence of Leu at P_C .

The C-terminal part of peptides bound to BF2*21:01 have very similar conformations

Thus far, eight structures have been determined for 10mer and 11mer peptides bound to the dominantly-expressed class I molecule of the B21 haplotype, BF2*21:01. Despite the very wide range of residues found for the anchor residues P_2 and P_{C-2} , as well as the non-anchor residues P_3 and P_{C-3} , the main chains of these eight peptides are nearly superimposable for the seven amino acids at the C-terminus (P_{C-4} through P_C), including the descending helical region of P_{C-5} through P_{C-3} (Fig. 28 [↗](#)). In contrast, the conformations in the N-terminal end of the peptide (P_2 and P_3 for 10mers, P_2 through P_4 for 11mers, and perhaps including P_1) vary significantly. B factors are also higher for the main chain atoms of the N-terminal part of the peptide compared to the C-terminal part, signifying greater flexibility in the structures.

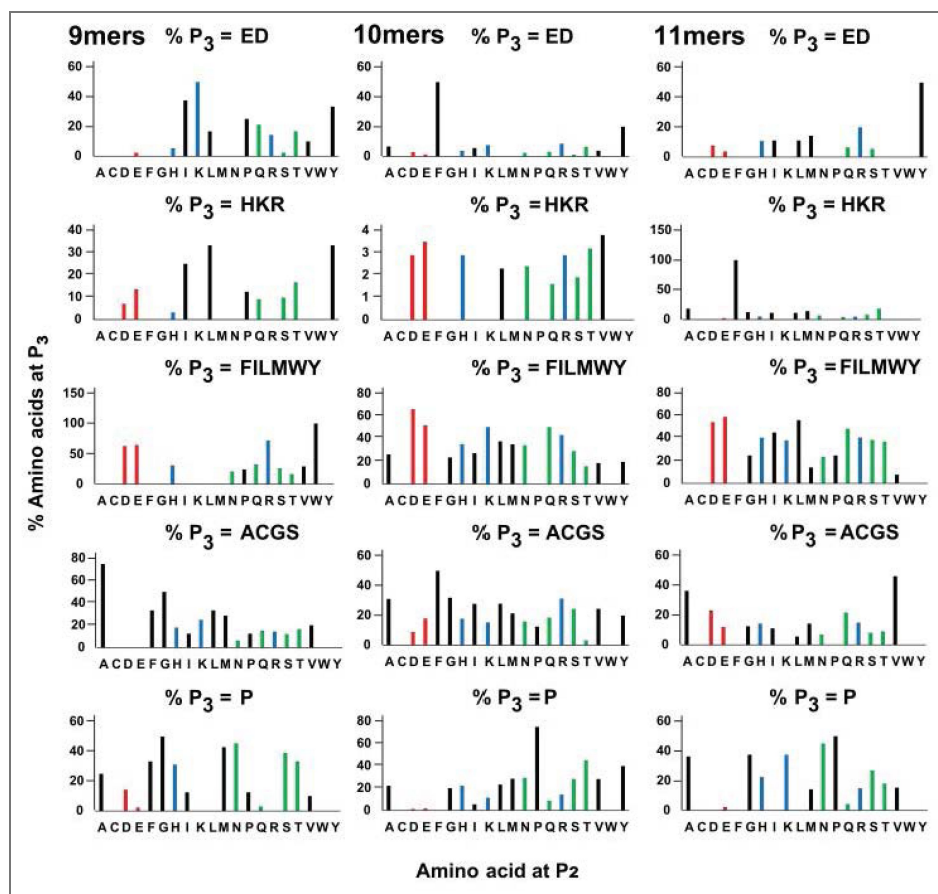


Figure 24. The frequencies of groups of amino acids (acidic, basic, large hydrophobic, small and Pro) found at position P_3 in peptides bound to class I molecules from a B21 cell line vary depending on the amino acid at P_2 and on the length of peptide.

Peptides identified by immunopeptidomics as in Fig. 17, with bar graphs showing the percentage of sequences with amino acids at P_2 on the x-axis against different groups of amino acids (single letter code: acidic being Asp and Glu; basic being His, Lys and Arg; large hydrophobic being Phe, Ile, Leu, Met, Trp and Tyr; small being Ala, Cys, Gly Ser; Pro) at P_3 on the y-axis, as found by mass spectroscopy.

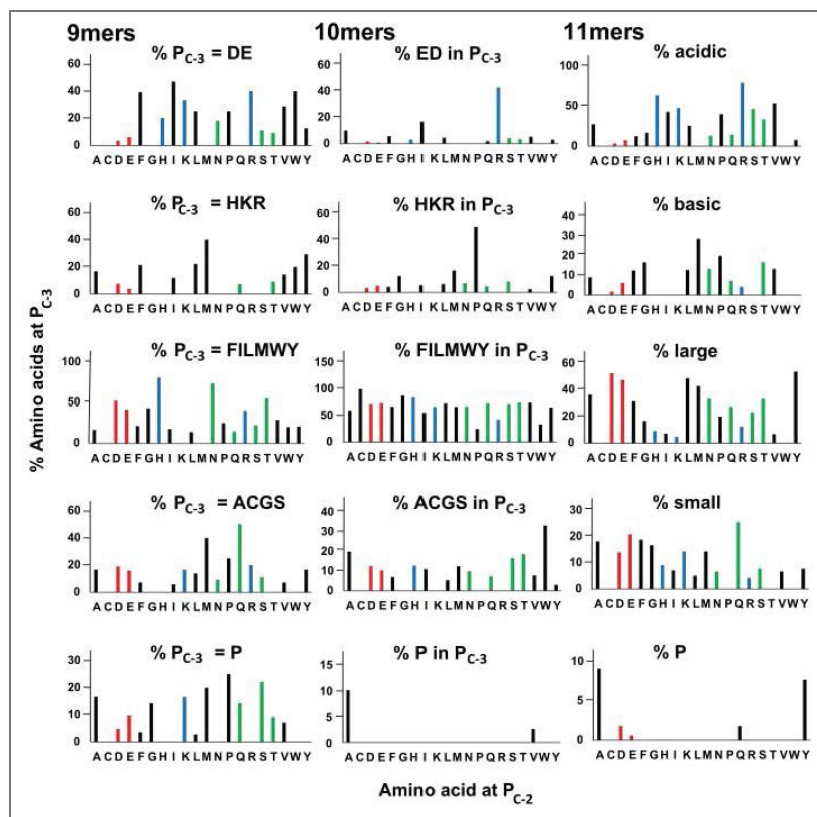


Figure 25. The frequencies of groups of amino acids (acidic, basic, large hydrophobic, small and Pro) found at positions P_{C-3} in peptides bound to class I molecules from a B21 cell line vary depending on the amino acid at P_{C-2} and on length of peptide, with 8mers more like 10mers.

Peptides identified by immunopeptidomics as in Fig. 17, with bar graphs showing the percentage of sequences with amino acids at P_{C-2} on the x-axis against different groups of amino acids (single letter code: acidic being Asp and Gly; basic being His, Lys and Arg; large hydrophobic being Phe, Ile, Leu, Met, Trp and Tyr; small being Ala, Cys, Gly Ser; Pro) at P_{C-3} on the y-axis, as found by mass spectroscopy.

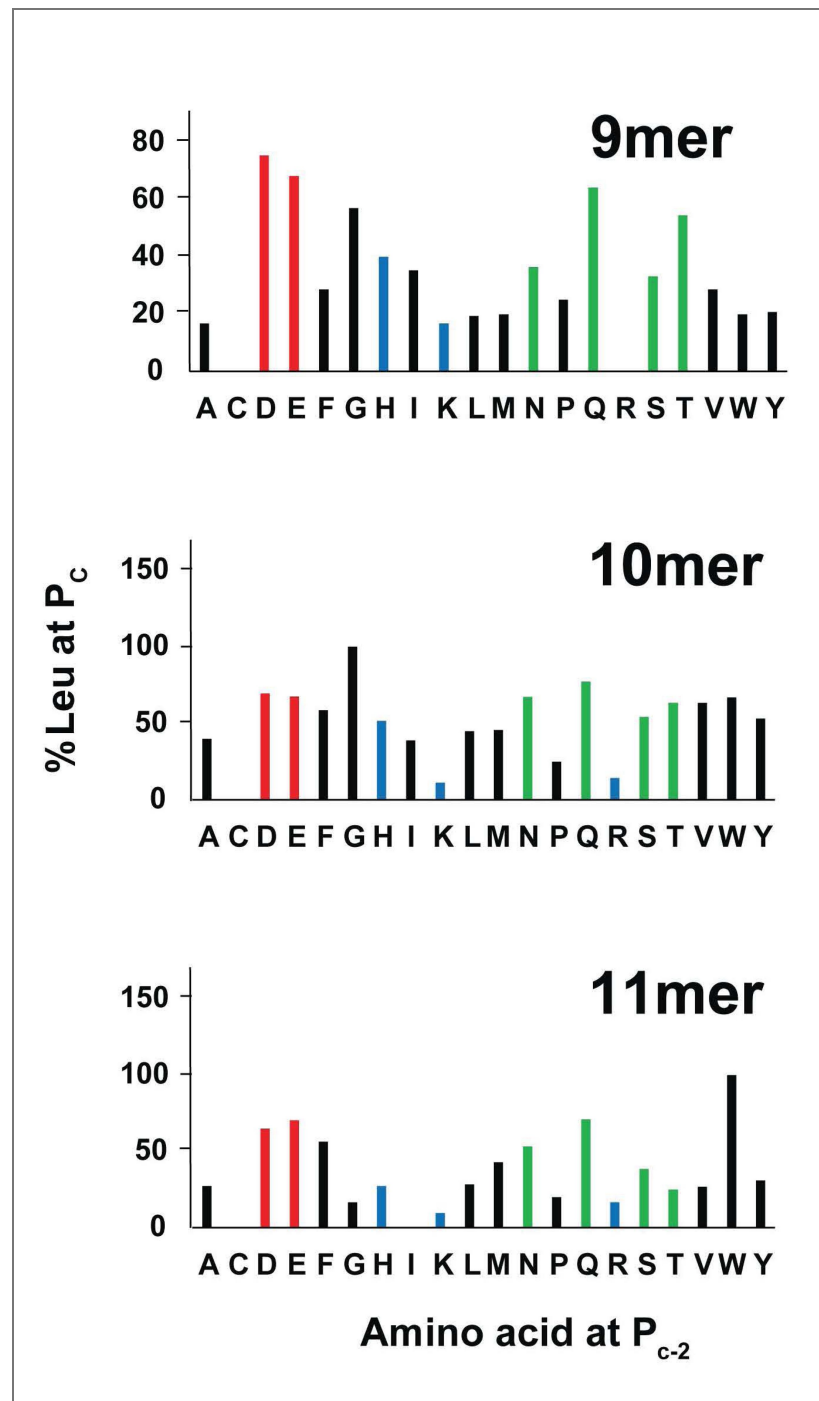


Figure 26. The frequency of Leu found at position P_c in peptides bound to class I molecules from a B21 cell line vary depending on length of peptide and position P_{c-2} .

Peptides identified by immunopeptidomics as in Fig. 17, with bar graphs showing the percentage of sequences with Leu at P_{c-2} on the y-axis against amino acids (single letter code) at P_c on the x-axis, as found by mass spectroscopy.

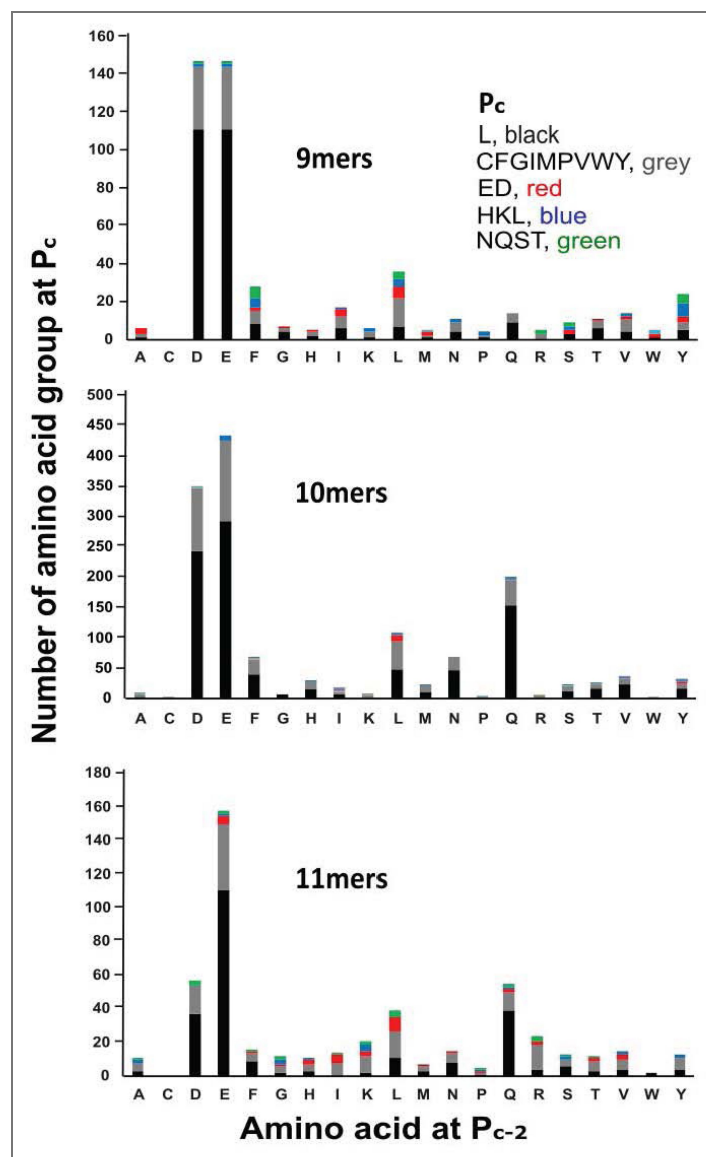


Figure 27. The frequencies of groups of amino acids (Leu, other hydrophobic, acidic, basic, polar) found at P_c in peptides bound to class I molecules from a B21 cell line vary depending on the amino acid at P_{c-2} and on length of peptide, with significant proportions of acidic, basic and/or polar amino acids compared to Leu at P_c for certain amino acids at P_{c-2} in 9mers and 11mers.

Peptides identified by immunopeptidomics as in Fig. 17, with stacked bar graphs showing the number of sequences with Leu, other hydrophobic (Cys, Phe, Gly, Ile, Met, Pro, Val, Trp and Tyr), acidic (Asp and Glu), basic (His, Lys and Arg) and polar (Asn, Gln, Ser and Thr) amino acids at P_c against length of peptide and amino acid (single letter code) at P_{c-2}, as found by mass spectroscopy.

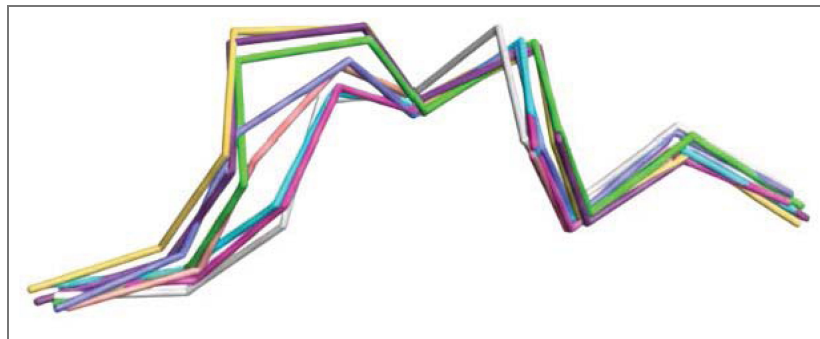


Figure 28. Superimposition of peptides from eight structures of BF2*21:01 show that the C-terminal portions all have similar conformations, while the N-terminal portions vary considerably.

Bacterially-expressed heavy chain and β_2 -microglobulin were refolded in vitro with various synthetic peptides, purified by SEC, crystallized and the X-ray structures determined. Depicted are the main chains, with N-termini on the left, viewed from side of α_2 helix of the class I molecule. Structures are 3BEV (GHAEYGAETL, green) and 3BEV (REVDEQLLSV, light blue) (Koch et al., 2008), 2YEZ (TNPEKVFYL, pink), 4D0B (TAGQEDYDRL, yellow), 4D0C (TAGQSNYDRL, wheat) and 4CVZ (YELDEKFDRL, grey) (Chappell et al., 2015 [DOI](#)), and 5AD0 (GHAEYGADTL, dark blue) and 5ACZ (GRAEYGADTL, orange) (this publication).

One interpretation of this finding is that, during peptide binding, the C-terminus sits down first in the peptide-binding groove, with the N-terminus then accommodating as best it can. This idea fits with the importance of the C-terminus in peptide transport by the TAPs correlated with peptide binding to class I molecules (Deverson et al., 1998 [↗](#); Joly et al., 1998 [↗](#); Momburg et al., 1994 [↗](#)) and with the action of ERAPs on the N-termini of bound but extended peptides (Li et al., 2019 [↗](#); Papakyriakou et al., 2018 [↗](#)), but may not fit as well with other interpretations (Mavridis et al., 2021 [↗](#)).

Discussion

The dominantly-expressed class I molecule of the chicken MHC haplotype B21, BF2*21:01, remodels the binding site to bind very different peptides, with co-variation of the anchor residues at P₂ and P_{C-2}, so that no simple peptide-binding motif was obvious (Chappell et al., 2015 [↗](#); Koch et al., 2008). To better understand the rules of peptide binding, we used X-ray crystal structures, molecular modelling and thermostability assays, as well as assembly assays for single peptides, single-substitution libraries and double-substitution libraries assessed by SEC, by SEC followed by MALDI-TOF and by SEC followed by LC-MS/MS. In addition, we used immunopeptidomics (immunoaffinity chromatography followed by LC-MS/MS) of class I molecules from a B21 cell line. The overall conclusion of this study is that the rules of peptide binding to BF2*21:01 are complex, but a few simplifying principles contribute to the relatively simple overall outcome.

As previously inferred from structures of BF2*21:01 with six different peptides originally identified from a cell line, blood cells and spleen cells (Chappell et al., 2015 [↗](#); Koch et al., 2008), both the assembly assays and immunopeptidomics show co-variation of anchor residues at P₂ and P_{C-2}. The complexity of binding was enormous: out of 400 possible combinations at these two positions, 159 were found in double-substitution libraries of four peptide backbones assessed by SEC followed by LC-MS/MS and 260 were found by a single immunopeptidomics experiment with one cell line. Given that a single sampling will miss many rare sequences in immunopeptidomics (discussed below), it seems likely that most or all combinations of amino acids at P₂ and P_{C-2} will turn out to be possible. This large number of combinations at P₂ and P_{C-2} underscores the plasticity of binding allowed by BF2*21:01 remodelling the peptide-binding site.

However, it appears that the stability of binding is affected by length of peptide. Based on assembly assays, thermostability and number of peptides found by immunopeptidomics, 10mers appear to be the optimal length, with 9mers and 11mers tolerated if optimal amino acids at P₂, P_{C-2} and P_C are present. There were some longer peptides found, but at much lower frequencies; these may bulge in the middle or hang out the C-terminal end (Burrows et al., 2006 [↗](#); Liu et al., 2012 [↗](#); Xiao et al., 2018 [↗](#)).

The stability of binding is affected also by the particular residues present at P₂ and P_{C-2}, including their combinations. Not all amino acids are optimal, with His at P₂ in the vast majority of 11mers, as assessed by immunopeptidomics. By assembly assay of peptide libraries, even the extremely stable 10mer REVDEQLLSV prefers His at P₂. The widest range at P₂ found for 10mers, as though they are able to tolerate less optimal amino acids. Even for 10mers, about a third of sequences each were found for Glu, for His and for the polar amino acids Asn, Gln and Ser, with the remainder mostly hydrophobic. Peptides with acidic residues at P₂ vastly preferred Gln at P_{C-2}, while basic and polar residues at P₂ vastly preferred Asp and Glu at P_{C-2}. These preferences were reflected in the optimal combinations for P₂ and P_{C-2} by immunopeptidomics, with the most frequent combination for 11mers accounting for only 12% of total sequences, and with the most combinations and the most even distribution for 10mers. Many of these combinations were shared among 9mers, 10mers and 11mers, as though such anchor residues confer the greatest stability at all three lengths.

In addition, the amino acids present at P₃ and P_{C-3} affect the anchor residues present at P₂ and P_{C-2}, further reducing the number of peptides that could be bound to BF2*21:01. The presence of particular residues at P₃ affects the stability of peptide, as assessed by assembly and

thermostability assays. Similarly, the combinations of P_2 and P_3 or of P_{c-2} and P_{c-3} are quite skewed as assessed by immunopeptidomics.

Finally, several amino acids are found at the anchor residue P_c by assembly assays with single peptides and with single-substitution libraries, but by immunopeptidomics far and away the most frequent was Leu. Assembly assays suggest that Ala and Trp are not favoured at P_c in the context of certain single peptides; one 11mer isolated from cells that had Trp at P_c had low thermostability, which was much improved by substitution with Leu. Assembly assays with peptide libraries based on the original 10mer and 11mer show that Leu/Ile, Phe, Met and Val are favoured. The polymorphic TAPs in chickens determine the peptides supplied to the class I molecules (Tregaskes et al., 2016 [↗](#); Walker et al., 2011 [↗](#)), so it seems likely that the overwhelming Leu at P_c by immunopeptidomics is imposed by the translocation specificity in B21 cells, although translocation assays with single synthetic peptides show that other residues can be accommodated (Tregaskes et al., 2016 [↗](#)). Whatever the reason, it is a further simplification reducing the number of peptides that can be bound to BF2*21:01 in vivo.

Confidence in the details of these interpretations must be tempered by the fact that both the assembly assays and immunopeptidomics have weaknesses, and that the two approaches do not give entirely the same results. Both MALDI-TOF and LC-MS/MS suffer from the weakness that peptides fly with different efficiencies, varying over orders of magnitude (Abdul-Khalek et al., 2025 [↗](#)). In that sense, the peak height, ion current and number of reads are not quantitative representations of the amount of a given peptide present. For the assembly assays, it seems very likely the results for any particular position or combination of position will depend on the other amino acids present, the “backbone” of the particular peptide, so a wide range of peptide backbones must be assessed. Moreover, assembly assays assess the ability of a peptide to refold in vitro, which can differ significantly from the immunopeptidomics results from cells for which various quality control mechanisms are important. It is already known for another chicken class I allele, BF2*15:01, that refolding in vitro allows amino acids at anchor residues that are never found on the surface of cells by gas phase sequencing (Tregaskes et al., 2016 [↗](#); Wallny et al., 2006 [↗](#)). Neither of these problems are so important for immunopeptidomics that can sample thousands of peptides simultaneously, but only a portion of very complex mixtures is sampled by the mass spectrometer, so both abundant and some of the rare peptides will be detected in each analysis. Therefore, the best way to determine whether a particular peptide is abundant is to detect it in many technical or biological replicates (for instance, Gastaldello et al., 2021 [↗](#); Halabi et al., 2021 [↗](#)). In addition, the potential for contribution by the poorly-expressed class I molecule, BF1*21:01, to the immunopeptidomic analysis is a concern, although BF1 molecules are generally expressed at very low levels compared to BF2 molecules (Wallny et al., 2006 [↗](#)). Also, only one biological replicate (with at least two technical replicates using different mass spec instruments) of a single cell line was assessed, and details may differ between cell types. The fact that results by the two approaches differed for P_c inspires further caution, particularly whether assembly assays with random libraries (for example, Jia et al., 2024 [↗](#); Li et al., 2024 [↗](#); Ma et al., 2020 [↗](#); Qu et al., 2019 [↗](#); Wei et al., 2021 [↗](#)) accurately reflect biological situations (Tregaskes et al., 2016 [↗](#)). Finally, chicken class I molecules allow peptides to extend out the C-terminal end of the peptide-binding groove (Xiao et al., 2018 [↗](#)), so the assignment of P_c (and therefore P_{c-2} and P_{c-3}) are not completely certain. The percentages of such C-terminal extensions are low for class I molecules with more easily understood peptide motifs (Han et al., 2023 [↗](#); Wallny et al., 2006 [↗](#)), but only experimental approaches like crystal structures can resolve this issue for sure.

These concerns about the details of the interpretation aside, we have determined much about peptide-binding to the dominantly-expressed class I molecule of the B21 haplotype, an MHC haplotype associated with resistance to a variety of economically-important diseases. By in vitro assembly assays and from living cells, BF2*21:01 has the capacity to bind most combinations of the two anchor residues P_2 and P_{c-2} , making it by far the most promiscuous class I molecule so far described. However, the most frequent and stable peptides have strong preferences for length, and for identity of the amino acids found at P_2 and P_{c-2} , as well as backbone positions such as P_3 and P_{c-3} . In addition, although four amino acids at P_c allow assembly in vitro, only Leu at P_c is

found with high frequency from the cell line. Some of these preferences can be understood by the structures and models determined in the course of this work. Taking these simplifying principles into account would allow prediction of peptides that will bind BF2*21:01 to better understand responses to infectious pathogens and to develop better poultry vaccines, but machine learning by artificial neural networks (Aranha et al., 2020 [↗](#); Jørgensen et al., 2014 [↗](#); Soam et al., 2011 [↗](#)) may allow even better predictions. Having said that, at the moment predictions by machine learning can give answers, but without true knowledge, which is the understanding of why. The results and interpretations in this study are the first step towards that understanding.

Materials and Methods

Key Resources Table				
Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
cell line (<i>Gallus gallus</i>)	TG21	PMID: 21536896 (Walker et al. 2011, PNAS)	In-house cell line	TG21 was derived by Dr. Thomas Göbel at the Basel Institute for Immunology in the early 1990s using reticuloendotheliosis virus REV-T transformation of concanavalin A-activated splenocytes from H.B21 birds.
cell line (<i>Gallus gallus</i>)	AVOL-1	PMID: 18256158 (Yao et al. 2008, J Virol)	In-house cell line	AVOL-1 was derived in the lab of Dr. Venugopal Nair at the Institute for Animal Health at Compton UK (now the Pirbright Institute) in the early 2000s using REV-T transformation of spleen cells from a line 0 chicken.

biological sample (<i>Gallus gallus</i>)	H.B21 chicken line	PMID: 6763913 (Simonsen et al. 1982, Immunogenetics); ISBN 0849367670, 9780849367670 (Hala K 1987 in: Toivanen A, Toivanen P, eds. Avian immunology: basis and practice. CRC, Boca Raton, pp 85–99); PMID: 15257423 (Miller et al. 2004, Immunogenetics)	In-house chicken line	The H.B21 chicken line was bred from American white leghorns in Copenhagen in the 1970s, and kept at the Basel Institute for Immunology until its closure.
biological sample (<i>Gallus gallus</i>)	N chicken line	PMID: 6763913 (Simonsen et al. 1982, Immunogenetics); ISBN 0849367670, 9780849367670 (Hala K 1987 in: Toivanen A, Toivanen P, eds. Avian immunology: basis and practice. CRC, Boca Raton, pp 85–99); PMID: 15257423 (Miller et al. 2004, Immunogenetics)	In-house chicken line	The N chicken line was bred from Hy-Line commercial birds by Randy Cole at Cornell University in 1965, then kept at the Regional Poultry Research laboratory in East Lansing Michigan, sent to the Institute for Animal Health at Compton UK and finally transferred to the Roslin Institute in Edinburgh Scotland, before being retained only as primordial germ cells.

antibody	F21-2 (Mouse monoclonal)	PMID: 3980052 (Crone et al. 1985 Immunogenetics), currently available from Southern Biotech (and Thermo-Fisher, Fisher Scientific, BioRad, Invitrogen, Arigo, etc)	Southern Biotech Cat. No. 8345-01	Monoclonal antibody F21-2 was raised in the lab of Morten Simonsen in Copenhagen against chicken cells and reacts with chicken class I molecules/heavy chains on cells, by immunoprecipitation and by western blot.
antibody	F21-21 (Mouse monoclonal)	PMID: 3980052 (Crone et al. 1985 Immunogenetics), currently available from Southern Biotech (and Thermo-Fisher, Fisher Scientific, BioRad, invitrogen, Arigo, etc)	Southern Biotech Cat. No. 8355-01	Monoclonal antibody F21-21 was raised in the lab of Morten Simonsen in Copenhagen against chicken cells and reacts with chicken class I molecules/beta2-microglobulin on cells, by immunoprecipitation and by western blot.

Cells and peptides. Blood was taken from H.B21 birds at the Gipf (Oberfrick) Farm of the Basel Institute for Immunology and processed to give peptides from class I molecules as described (Wallny et al., 2006 [\[link\]](#)). Spleens were taken from N line birds at the Institute for Animal Health (Compton, UK) after retirement from breeding flock and schedule 1 procedures, and processed to give peptides from class I molecules as described (Chappell et al., 2015 [\[link\]](#)). The B21 cell line TG21 was derived by reticuloendotheliosis virus REV-T transformation of concanavalin A-activated splenocytes from H.B21 birds at the Basel Institute for Immunology and maintained as described (Walker et al., 2011 [\[link\]](#)), with peptides isolated from class I molecules as described (Chappell et al., 2015 [\[link\]](#)). The peptides from blood, spleen and TG12 cells were separated by HPLC and single peaks subjected to gas phase sequencing as described (Chappell et al., 2015 [\[link\]](#); Wallny et al., 2006 [\[link\]](#)). The B21 cell line AVOL-1 was derived by REV-T transformation of spleen cells from a line 0 chicken at the Pirbright Institute, with peptides isolated from class I molecules and prepared for immunopeptidomics, all as described (Chappell et al., 2015 [\[link\]](#)).

Expression and isolation of protein chains. The cloning of BF2*21:01 and β_2 -microglobulin sequences, recloning into pET22b(+) expression plasmids, expression in BL21 (λ DE3) pLysS Rosetta cells, isolation, washing and solubilisation of inclusion bodies using urea were broadly as previously described (Chappell et al., 2015 [\[link\]](#); Koch et al., 2007 [\[link\]](#)), but occasionally used TopTen medium, which led to much higher yields.

Refolding with peptides. Renaturation of BF2*21:01 heavy chain with β_2 -microglobulin and synthetic peptides, separation by size exclusion chromatography and concentration of monomer fractions were broadly as previously described (Chappell et al., 2015 [DOI](#); Koch et al., 2007 [DOI](#)). Single peptides and peptide libraries were synthesised by fmoc chemistry and appropriately purified (Pepceuticals), with lyophilised material dissolved in 0.1 % trifluoroacetic acid. The peptide libraries had one or two positions at which all natural amino acids except cysteine were coupled, using ratios of reactants previously determined to give more-or-less equivalent incorporation. The renaturation buffer used was 400 mM arginine, 2 mM ethylene diamine tetraacetic acid (EDTA), 100 mM TrisCl pH 8.2, with oxidized glutathione added to 0.5 mM, reduced glutathione added to 5 mM and 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF) added to 0.5 mM just before use. The refolding involved adding denatured heavy chain with rapid stirring to refolded β_2 -microglobulin and synthetic peptide(s) in renaturation buffer, and incubating with continued stirring 18-40 hours at 4°C. The samples were centrifuged at high speed to precipitate flocculate material, and the supernatant was passed through 0.2 or 0.45 micrometer sterile filters before separation by size exclusion chromatography in TBS (100 mM NaCl, 25 mM TrisCl pH 8.0) using an AKTA 920 running at 1 ml/min at room temperature, followed by concentration of fractions corresponding to refolded monomers.

The total volume of refold mixture was typically 100 ml for single peptide refolds intended for crystallization and for thermostability assays, and 1 ml or 1.8 ml for all other refolds involving peptide libraries. Most refolding was at molar concentrations of roughly 1 mole heavy chain: 2 mole β_2 -microglobulin: 10 mole peptide (1:2:10), except for some analytical runs with peptide libraries in which 1:2:10 was compared with 1:2:2 (the latter to control for one peptide outcompeting other peptides with slightly lower affinity). Heavy chain was added slowly dropwise to rapidly stirred refold mix containing peptide and previously refolded β -microglobulin. After refolding at 4°C, some samples were analysed immediately, while others were incubated at 4°C or at 42°C (the latter to assess whether the complexes stable at 4°C might unfold at the body temperature typical of a chicken) for 1, 6 or 18 hours (results shown for 18 hours in Fig. 8 [DOI](#)).

The larger scale refolds were chromatographed on a HiLoad 26/60 Superdex 200 column (GE Healthcare, UK) column at room temperature. Fractions corresponding to monomer were pooled and concentrated with a Vivaspin 50 (10 kDa cut-off, Sartorius) at 4°C until the volume was a few ml. Then, a Vivaspin 2 (10 kDa cut-off, Sartorius) at 4°C was used to replace the buffer with 50 mM NaCl, 10 mM TrisCl pH 8 and to reduce the volume to 100-200 microliters, which were snap frozen on dry ice before storage at -80°C until use.

The smaller scale refolds were chromatographed on a Superdex 200 10/300 GL column (GE Healthcare, UK) column, with fractions corresponding to monomer pooled and concentrated first using a pre-rinsed Amicon Ultra-4 10 kDa column and then a Vivaspin 500 10 kDa column to 100-200 microliters, and the buffer exchanged to 10 mM TrisCl pH 8. The sample was further concentrated using a SpeedVac at 40°C to 5-10 microliters, glacial acetic acid was added to 5% v/v followed by further concentration by SpeedVac at 40°C to 2-4 microliters, before freezing at -80°C. Analysis was typically by matrix-assisted laser-desorption ionisation-time of flight mass spectroscopy (MALDI-TOF, Protein and Nucleic Acid Chemistry services, Department of Biochemistry, University of Cambridge), but later by liquid chromatography-mass spectroscopy/mass spectroscopy (LC-MS/MS, Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge). For MALDI-TOF, spectra were analysed by comparison to mass-to-charge (m/z) ratio calculated for each peptide in the library, with the amount of each peptide detected in the input mixture verified by MALDI-TOF. For LC-MS/MS, the peptides were identified using Mascot. Calculations, comparisons and graph-making were performed using Excel, and figures made using Inkscape.

Thermostability assays. Refolds of 6 ml were performed and split into 1 ml aliquots at 1 mg/ml. The thermal stabilities of BF2*21:01 with different peptides were assessed using the ThermoFluor assay (Matulis et al., 2005 [DOI](#)). In detail, 20 μ L of each protein solution at 1 mg/mL was added to 5 μ L of SPYRO orange dye (Life Technologies) (1000-fold dilution) in a 96-well plate. The temperature

was increased at a rate of 1.5°C per minute and fluorescence monitored using a Stratagene MX3005P real time PCR machine. Fluorescence measurements were normalised and then fitted to a Boltzmann sigmoidal curve using GraphPad Prism to calculate melting temperatures.

Crystallography. Protein crystals were obtained and structures were determined overall as described (Chappell et al., 2015 [DOI](#)). In detail, recombinant MHC class I complexes were crystallized using the sitting-drop method [0.2 M MgCl₂·(H₂O)₆, 0.1 M Tris, pH 8.0, 20% PEG6000 for 5AD0 with peptide GHAEYEGADTL; Molecular Dimensions E9, pH 8.5 for 5ACZ with peptide GRAEYEGADTL]. Crystals were flash frozen in liquid nitrogen and native data sets for each crystal were collected at 100K [ESRF, Grenoble (beamline ID23-2) for 5AD0; Diamond Light Source, Harwell (beamline I02) for 5ACZ]. Data were processed using either AUTOPROC (Vonnrhein et al., 2011 [DOI](#)) or XIA2 (Winter, 2010 [DOI](#)) with XDS (Kabsch, 2010 [DOI](#)) for integration and AIMLESS (Evans, 2006 [DOI](#)) or SCALA (Evans and McCoy, 2008 [DOI](#)) for scaling. All structures were solved by molecular replacement, as implemented in Phaser (McCoy et al., 2007 [DOI](#)), part of the CCP4 software package (Winn et al., 2011 [DOI](#)). Starting molecular replacement models were generated using CHAINSAW (Stein, 2008 [DOI](#)) and the atomic co-ordinates of the chicken B21 MHC class I molecule (PDBID: 3BEV) with the peptide removed. Model building and refinement were carried out using COOT (Emsley et al., 2010 [DOI](#)) and AUTOBUSTER (Bricogne et al., 2011 [DOI](#)) or REFMAC (Murshudov et al., 1997 [DOI](#)), with the heavy and light chains of the MHC molecule rebuilt first before the peptide was modelled into residual electron density. Structures were depicted using PyMol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

Modelling. Beginning with known crystal structures, modelling was carried out using PyMol with the protein mutagenesis Wizard, the rotamer toggle, show bumps and show surfaces.

Immuno-peptidomics. Peptide sequences from AVOL-1 cells were determined just as described (Chappell et al., 2015 [DOI](#)). In detail, immunoaffinity beads were produced, with all steps at room temperature. Protein G-Sepharose beads (Expediton) were washed with 50 mM borate, 50 mM KCl, pH 8.0, the equivalent of 1-ml packed beads was incubated with 3 mg F21-2 (produced by the Microbiological Media Services of the Pirbright Institute) for 1 hr, treated with 40 mM dimethyl pimelimidate dihydrochloride (Sigma) in 0.1 M triethanolamine, pH 8.3 for 1 hr to cross-link the antibody to the protein G, washed with 100 mM citric acid pH 3.0, and equilibrated in 50 mM TrisHCl, pH 8.0. The cell line AVOL-1 was washed with PBS. A pellet of 10⁹ cells was incubated with 10 ml 1% Igepal 630, 300 mM NaCl, 100 mM Tris pH 8.0 for 30 min at 4°C, subcellular debris was pelleted by centrifugation at 300×g for 10 min and 15,000×g for 30 min at 4°C, and the cleared lysates were incubated with 1 ml immunoaffinity beads for 1 hr at 4°C. The beads were washed with 50 mM TrisCl, pH 8.0 buffer, first with 150 mM NaCl, then with 400 mM NaCl and finally with no salt.

Bound material was eluted with 10% acetic acid. The eluted material was dried, resuspended in 3% acetonitrile, 0.1% formic acid in water, loaded directly onto on a 4.6 × 50 mm ProSwift™ RP-1S column (ThermoFisher) and eluted at 500 µl/min flow rate for 10 min with a linear gradient from 2 to 35% buffer B (0.1% formic acid in acetonitrile) in buffer A (0.1% formic acid in water) using an Ultimate 3000 HPLC system (ThermoFisher), with fractions collected from 2 to 15 min. Protein detection was performed at 280 nm absorbance, with fractions eluting before β₂-microglobulin pooled and dried.

For liquid chromatography tandem mass spectrometry (LC-MS/MS), peptides were analysed using either a Q-Exactive (Thermo Scientific) or a TripleTOF 5600 (AB Sciex) system, as identified (Dataset 1). For the QExactive system, peptides were separated on a Ultimate 3000 RSLCnano System utilizing a PepMap C18 column, 2 µm particle size, 75 µm × 50 cm (Thermo Scientific) with a linear gradient from 3% to 35% buffer B in buffer A (as above) at a flow rate of 250 nl/min (~ 65 MPa) for 60 min, and the 15 most intense precursors per full MS scan were selected for MS/MS analysis using HCD fragmentation. For the TripleTOF system, peptides were separated with a 15 cm × 75 µm ChromXP C18-CL (3 µm particle size) using an ekspert nanoLC 400 cHiPLC system (Eksigent) with a linear gradient from 8% to 35% buffer B in buffer A (as above) at a flow rate of 300 nl/min (~1600 psi) for 60 min, and CID fragmentation was induced on the 30 most abundant ions per full MS scan. All fragmented precursor ions were actively excluded from repeated

selection for 15 s. Data were analysed using Peaks 7 (Bioinformatics Solutions) using a database containing all 24,092 Uniprot entries for the organism *Gallus gallus* combined with protein translations (>8 amino acids) of either all six reading frames of gallid herpesvirus 2 (NCBI entry NC_002229.3; 10,026 entries) or reticuloendotheliosis virus (NC_006934.1; 412 entries). Results were filtered using a false discovery rate of 1% that was determined by parallel searching of a randomized decoy database. Peptide motifs were determined by Gibbs clustering (Andreatta et al., 2013) through a website (<https://services.healthtech.dtu.dk/services/GibbsCluster-2.0/>).

Table 1.

BF2*21:01	5ACZ	5AD0
Data Collection		
Wavelength (Å)	0.97	0.8726
Resolution range (Å) (highest resolution shell)	51.31 - 2.69 (2.82 - 2.69)	73.78- 2.84 (3.17- 2.84)
Space group	P 21 21 21	P 21 21 21
Unit cell a,b,c, (Å) a,b,g (°)	71.61, 72.33, 72.81 90, 90, 90	73.86, 72.33, 73.53 90.00, 90.00, 90.00
Total reflections	57185	54726
Unique reflections	10877	9390
Multiplicity	5.3	5.8
Completeness (%)	99.4 (97.7)	96.5 (87.7)
Mean I/sigma (I)	12.0 (2.7)	11 (2)
Wilson B-factor (Å ²)	33.5	41
R-merge	0.15 (0.69)	0.16 (0.41)
R-meas	0.267 (0.239)	0.250 (0.290)
R-pim	0.15 (0.69)	0.16 (0.41)
correlation coefficient Fo-Fc	0.894	0.869
Refinement		
Reflections used in refinement	10377	9355
Reflections used for R-free	474 (4.37%)	449 (4.8 %)
R _{work}	0.239	0.236
R _{free}	0.267	0.273
Number of atoms:	3031	3063
non-hydrogen		
macromolecules	0	0
Ligands	4	9
Solvent	4	32
Protein residues	3023	3022
RMSD (bonds, Å)	0.007	0.007
RMSD (angles, °)	1.1	0.8
Ramachandran favored (%)	98	96
Ramachandran allowed (%)	2	4
Ramachandran outliers (%)	0	0
Clash score	2	1
Average B-factor (Å ²)	30.0	38.0
Macromolecules	31	39
Ligands	35	40
Solvent	24	18
Number of TLS groups	0	0

Table 1.

Data availability

Diffraction data have been deposited in PDB under the accession code s 5ACZ and 5AD0. Harrison et al Supplemental data - Dataset1 contains the numerical data used to generate figures 10-24.

Acknowledgements

many thanks to researchers for help to identify the first peptides (Fig. 1 [↗](#)) at the Basel Institute for Immunology including Jan Salomonsen, and at the IAH Compton including Lawrence Hunt, Fiona Johnston, Denise A. Marston, Andy van Hateren, Simon Camp, Jansen P. Jacob, Iain Shaw, Timothy J. Powell, Sally Rogers and Brian Walker. Also thanks to Pietro Roversi at Oxford for help with the structures, and Laura Mears, Xaquín Castro-Dopico and Len Packman for help with the peptide libraries. Funding at the time the research was being done included support from Hofmann La Roche at the Basel Institute for Immunology to JS, DWA, HJW and JK, and from the Biotechnology and Biological Sciences Research Council (BBSRC) at the Institute for Animal Health Compton (later renamed the Pirbright Institute) to VN, WM and JK. In addition, grants included Wellcome Trust Studentships to LM and CH in Cambridge and a BBSRC studentship to PC at Oxford, Wellcome Trust Senior Investigator Award to SML at Oxford, grants to NT at Oxford, a program grant (089305/Z/09/Z) and an Investigator Award from the Wellcome Trust (110106/Z/15/Z) to JK in Cambridge and later in Edinburgh, and a BBSRC project grant (BB/V000756/1) and a German Science Foundation (DFG) consortium grant (FOR5130 - ImmunoChick, KA 5564/1) to JK in Edinburgh.

Additional information

Author contributions

Conception and design: VN, SML, NT, JK

Acquisition of data: MH, PEC, MJD, LM, CH, MJD, HJW, DWA, WM

Analysis and interpretation of data: MH, PEC, SH, MJD, KSL, HJW, DWA, WM, NT, SML, JK

Drafting or revising the article: EMM, SH, NT, SML, JK (all authors had the opportunity to read and comment on the article)

Ethics

Ethics involving chickens were carried out at the Basel Institute under the provisions in place at the time, and at the Institute for Animal Health under Home Office project license PPL 80/2420 and with ethical approval of the Local Ethical Review Committee.

Funding

Funder	Grant reference number	Author
Wellcome Trust (WT)	https://doi.org/10.35802/089305	Jim Kaufman
Wellcome Trust (WT)	https://doi.org/10.35802/110106	Jim Kaufman
UKRI Biotechnology and Biological Sciences Research Council (AFRC)	BB/V000756/1	Jim Kaufman
Deutsche Forschungsgemeinschaft (DFG)	KA 5564/1	Jim Kaufman

Author ORCID iDs

Clemens Hermann:  <https://orcid.org/0000-0002-0009-9501>

Susan M Lea:  <https://orcid.org/0000-0001-9287-8053>

Jim Kaufman:  <https://orcid.org/0000-0002-7216-8422>

Author notes

Samer Halabi: Current address: University of Edinburgh, Roslin Institute, Midlothian, United Kingdom

Łukasz Magiera: Current address: AstraZeneca, Oncology Targeted Discovery, The Discovery Centre, Cambridge Biomedical Campus, Cambridge, United Kingdom

Hans-Joachim Wallny: Current address: Wallny BiotechConsulting, Grenzach-Wyhlen, Germany

William Mwangi: Current address: LifeArc, Biologics Discovery & Development, The Francis Crick Institute, London, United Kingdom

Competing interests: No competing interests declared

Additional files

Figure 11—source data 1. [Raw data for ion current \(counts\) from MALDI-TOF and LC-MS/MS analyses of the peptides that assembled with BF2*21:01 from six double-substitution libraries \(source of data for bar graphs in Figs. 11 \[11\]\(#\)–\[13\]\(#\)\).](#)

Figure 17—source data 1. [Raw data from immunopeptidomics of class I molecules from AVOL-1 cells \(source of data for Figs. 17 \[17\]\(#\)–\[27\]\(#\)\).](#)

References

- Abdul-Khalek N,** Picciani M, Shouman O, Wimmer R, Overgaard MT, Wilhelm M, Gregersen Echers S (2025) To Fly, or Not to Fly, That Is the Question: A Deep Learning Model for Peptide Detectability Prediction in Mass Spectrometry. *J Proteome Res* **24**:2709–2726 <https://doi.org/10.1021/acs.jproteome.4c00973> | [PubMed](#) | [PubMed Central](#)
- Andreatta M,** Lund O, Nielsen M (2013) Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. *Bioinformatics* **29**:8–14 <https://doi.org/10.1093/bioinformatics/bts621> | [PubMed](#)
- Aranha MP,** Jewel YSM, Beckman RA, Weiner LM, Mitchell JC, Parks JM, Smith JC (2020) Combining Three-Dimensional Modeling with Artificial Intelligence to Increase Specificity and Precision in Peptide-MHC Binding Predictions. *J Immunol* **205**:1962–1977 <https://doi.org/10.4049/jimmunol.1900918> | [PubMed](#) | [PubMed Central](#)
- Bashirova AA,** Viard M, Naranbhai V, Grifoni A, Garcia-Beltran W, Akdag M, Yuki Y, Gao X, O’hUigin C, Raghavan M, *et al.* (2020) HLA tapasin independence: broader peptide repertoire and HIV control. *Proc Natl Acad Sci U S A* **117**:28232–28238 <https://doi.org/10.1073/pnas.2013554117> | [PubMed](#) | [PubMed Central](#)
- Bricogne G,** Blanc E, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi P, Sharff A, Smart OS, Vonnrhein C, *et al.* (2011) BUSTER.
- Burrows SR,** Rossjohn J, McCluskey J (2006) Have we cut ourselves too short in mapping CTL epitopes?. *Trends Immunol* **27**:11–6 <https://doi.org/10.1016/j.it.2005.11.001> | [PubMed](#)
- Chappell P,** Meziane el K, Harrison M, Magiera Ł, Hermann C, Mears L, Wrobel AG, Durant C, Nielsen LL, Buus S, *et al.* (2015) Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. *eLife* **4**:e05345 <https://doi.org/10.7554/eLife.05345> | [PubMed](#) | [PubMed Central](#)
- Crone M,** Simonsen M, Skjødt K, Linnet K, Olsson L (1985) Mouse monoclonal antibodies to class I and class II antigens of the chicken MHC. Evidence for at least two class I products of the B complex. *Immunogenetics* **21**:181–7 <https://doi.org/10.1007/BF00364870> | [PubMed](#)
- Djaoud Z** (2020) and KIRs, a Triumvirate of Human Cell-Mediated Immunity. *Annu Rev Biochem* **89**:717–739 <https://doi.org/10.1146/annurev-biochem-011520-102754> | [PubMed](#)
- Deverson EV,** Leong L, Seelig A, Coadwell WJ, Tredgett EM, Butcher GW, Howard JC (1998) Functional analysis by site-directed mutagenesis of the complex polymorphism in rat transporter associated with antigen processing. *J Immunol* **160**:2767–79 [PubMed](#)

- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**:486-501 <https://doi.org/10.1107/S0907444910007493> | PubMed | PubMed Central
- Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**:72-82 <https://doi.org/10.1107/S0907444905036693> | PubMed
- Evans P, McCoy A (2008) An introduction to molecular replacement. *Acta Crystallogr D Biol Crystallogr* **64**:1-10 <https://doi.org/10.1107/S0907444907051554> | PubMed | PubMed Central
- Gastaldello A, Ramarathinam SH, Bailey A, Owen R, Turner S, Kontouli N, Elliott T, Skipp P, Purcell AW, Siddle HV (2021) The immunopeptidomes of two transmissible cancers and their host have a common, dominant peptide motif. *Immunology* **163**:169-184 <https://doi.org/10.1111/imm.13307> | PubMed | PubMed Central
- Hala K (1987) Inbred lines of avian species. In: Toivanen A, Toivanen P (Eds). *Avian immunology: basis and practice* CRC, Boca Raton. pp. 85-99
- Halabi S, Kaufman J (2022) New vistas unfold: Chicken MHC molecules reveal unexpected ways to present peptides to the immune system. *Front Immunol* **13**:886672 <https://doi.org/10.3389/fimmu.2022.886672> | PubMed | PubMed Central
- Halabi S, Ghosh M, Stevanović S, Rammensee HG, Bertzbach LD, Kaufer BB, Moncrieffe MC, Kaspers B, Härtle S, Kaufman J (2021) The dominantly expressed class II molecule from a resistant MHC haplotype presents only a few Marek's disease virus peptides by using an unprecedented binding motif. *PLoS Biol* **19**:e3001057 <https://doi.org/10.1371/journal.pbio.3001057> | PubMed | PubMed Central
- Han L, Wu S, Zhang T, Peng W, Zhao M, Yue C, Wen W, Cai W, Li M, Wallny HJ, et al. (2023) A Wider and Deeper Peptide-Binding Groove for the Class I Molecules from B15 Compared with B19 Chickens Correlates with Relative Resistance to Marek's Disease. *J Immunol* **210**:668-680 <https://doi.org/10.4049/jimmunol.2200211> | PubMed | PubMed Central
- Jia Y, Wu Q, Li Y, Ma M, Song W, Chen R, Yao Y, Nair V, Zhang N, Liao M, et al. (2024) Revealing novel and conservative T-cell epitopes with MHC B2 restriction on H9N2 avian influenza virus (AIV). *J Biol Chem* **300**:107395 <https://doi.org/10.1016/j.jbc.2024.107395> | PubMed | PubMed Central
- Joly E, Le Rolle AF, González AL, Mehling B, Stevens J, Coadwell WJ, Hünig T, Howard JC, Butcher GW (1998) Co-evolution of rat TAP transporters and MHC class I RT1-A molecules. *Curr Biol* **8**:169-72 [https://doi.org/10.1016/s0960-9822\(98\)70065-x](https://doi.org/10.1016/s0960-9822(98)70065-x) | PubMed
- Jørgensen KW, Rasmussen M, Buus S, Nielsen M (2014) NetMHCstab - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology* **141**:18-26 <https://doi.org/10.1111/imm.12160> | PubMed | PubMed Central
- Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**:125-32 <https://doi.org/10.1107/S0907444909047337> | PubMed | PubMed Central
- Kaufman J (2015) Co-evolution with chicken class I genes. *Immunol Rev* **267**:56-71 <https://doi.org/10.1111/imr.12321> | PubMed
- Kaufman J (2018) Generalists and Specialists: A New View of How MHC Class I Molecules Fight Infectious Pathogens. *Trends Immunol* **39**:367-379 <https://doi.org/10.1016/j.it.2018.01.001> | PubMed | PubMed Central
- Kaufman J, Völk H, Wallny HJ (1995) A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. *Immunol Rev* **143**:63-88 <https://doi.org/10.1111/j.1600-065x.1995.tb00670.x> | PubMed
- Koch M, Camp S, Collen T, Avila D, Salomonsen J, Wallny HJ, van Hateren A, Hunt L, Jacob JP, Johnston F, et al. (2007) Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. *Immunity* **27**:885-99 <https://doi.org/10.1016/j.immuni.2007.11.007> | PubMed

- Li L, Batliwala M, Bouvier M (2019) ERAP1 enzyme-mediated trimming and structural analyses of MHC I-bound precursor peptides yield novel insights into antigen processing and presentation. *J Biol Chem* **294**:18534-18544 <https://doi.org/10.1074/jbc.RA119.010102> | PubMed | PubMed Central
- Li X, Li Z, Ma M, Yang N, Du S, Liao M, Dai M (2024) Revealing novel and conservative CD8⁺ T-cell epitopes with MHC B2 restriction on ALV-J. *Vet Res* **55**:164 <https://doi.org/10.1186/s13567-024-01426-3> | PubMed | PubMed Central
- Liu YC, Chen Z, Burrows SR, Purcell AW, McCluskey J, Rossjohn J, Gras S (2012) The energetic basis underpinning T-cell receptor recognition of a super-bulged peptide bound to a major histocompatibility complex class I molecule. *J Biol Chem* **287**:12267-76 <https://doi.org/10.1074/jbc.M112.344689> | PubMed | PubMed Central
- Ma L, Zhang N, Qu Z, Liang R, Zhang L, Zhang B, Meng G, Dijkstra JM, Li S, Xia MC (2020) A Glimpse of the Peptide Profile Presentation by *Xenopus laevis* MHC Class I: Crystal Structure of Xela-UAA Reveals a Distinct Peptide- Binding Groove. *J Immunol* **204**:147-158 <https://doi.org/10.4049/jimmunol.1900865> | PubMed | PubMed Central
- Manczinger M, Boross G, Kemény L, Müller V, Lenz TL, Papp B, Pál C (2019) Pathogen diversity drives the evolution of generalist MHC-II alleles in human populations. *PLoS Biol* **17**:e3000131 <https://doi.org/10.1371/journal.pbio.3000131> | PubMed | PubMed Central
- Matulis D, Kranz JK, Salemme FR, Todd MJ (2005) Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry* **44**:5258-66 <https://doi.org/10.1021/bi048135v> | PubMed
- Mavridis G, Mpakali A, Zoidakis J, Makridakis M, Vlahou A, Kaloumenou E, Ziotopoulou A, Georgiadis D, Papakyriakou A, Stratikos E (2021) The ERAP1 active site cannot productively access the N-terminus of antigenic peptide precursors stably bound onto MHC class I. *Sci Rep* **11**:16475 <https://doi.org/10.1038/s41598-021-95786-x> | PubMed
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr*. **40**:658-674 <https://doi.org/10.1107/S0021889807021206> | PubMed | PubMed Central
- Miller MM, Jr Taylor RL (2016) Brief review of the chicken Major Histocompatibility Complex: the genes, their distribution on chromosome 16, and their contributions to disease resistance. *Poult Sci* **95**:375-92 <https://doi.org/10.3382/ps/pev379> | PubMed | PubMed Central
- Miller MM, Bacon LD, Hala K, Hunt HD, Ewald SJ, Kaufman J, Zoorob R, Briles WE (2004) Nomenclature for the chicken major histocompatibility (B and Y) complex. *Immunogenetics* **56**:261-79 <https://doi.org/10.1007/s00251-004-0682-1> | PubMed
- Momburg F, Roelse J, Howard JC, Butcher GW, Hämmerling GJ, Neefjes JJ (1994) Selectivity of MHC-encoded peptide transporters from human, mouse and rat. *Nature* **367**:648-51 <https://doi.org/10.1038/367648a0> | PubMed
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**:240-55 <https://doi.org/10.1107/S0907444996012255> | PubMed
- Papakyriakou A, Reeves E, Beton M, Mikolajek H, Douglas L, Cooper G, Elliott T, Werner JM, James E (2018) The partial dissociation of MHC class I-bound peptides exposes their N terminus to trimming by endoplasmic reticulum aminopeptidase 1. *J Biol Chem* **293**:7538-7548 <https://doi.org/10.1074/jbc.RA117.000313> | PubMed | PubMed Central
- Paul S, Weiskopf D, Angelo MA, Sidney J, Peters B, Sette A (2013) HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. *J Immunol* **191**:5831-9 <https://doi.org/10.4049/jimmunol.1302101> | PubMed | PubMed Central
- Purcell AW, Ramarathinam SH, Ternette N (2019) Mass spectrometry-based identification of MHC-bound peptides for immuno-peptidomics. *Nat Protoc* **14**:1687-1707 <https://doi.org/10.1038/s41596-019-0133-y> | PubMed

- Qu Z, Li Z, Ma L, Wei X, Zhang L, Liang R, Meng G, Zhang N, Xia C (2019) Structure and Peptidome of the Bat MHC Class I Molecule Reveal a Novel Mechanism Leading to High-Affinity Peptide Binding. *J Immunol* **202**:3493-3506 <https://doi.org/10.4049/jimmunol.1900001> | PubMed | PubMed Central
- Rizvi SM, Salam N, Geng J, Qi Y, Bream JH, Duggal P, Hussain SK, Martinson J, Wolinsky SM, Carrington M, et al. (2014) Distinct assembly profiles of HLA-B molecules. *J Immunol* **192**:4967-76 <https://doi.org/10.4049/jimmunol.1301670> | PubMed | PubMed Central
- Shaw I, Powell TJ, Marston DA, Baker K, van Hateren A, Riegert P, Wiles MV, Milne S, Beck S, Kaufman J (2007) Different evolutionary histories of the two classical class I genes BF1 and BF2 illustrate drift and selection within the stable MHC haplotypes of chickens. *J Immunol* **178**:5744-52 <https://doi.org/10.4049/jimmunol.178.9.5744> | PubMed
- Simonsen M, Crone M, Koch C, Håla K (1982) The MHC haplotypes of the chicken. *Immunogenetics* **16**:513-32 <https://doi.org/10.1007/BF00372021> | PubMed
- Soam SS, Bhasker B, Mishra BN (2011) Improved prediction of MHC class I binders/non-binders peptides through artificial neural network using variable learning rate: SARS corona virus, a case study. *Adv Exp Med Biol* **696**:223-9 https://doi.org/10.1007/978-1-4419-7046-6_22 | PubMed | PubMed Central
- Stein N (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography* **41**:641-643 <https://doi.org/10.1107/S0021889808006985>
- Tregaskes CA, Kaufman J (2021) Chickens as a simple system for scientific discovery: The example of the MHC. *Mol Immunol* **135**:12-20 <https://doi.org/10.1016/j.molimm.2021.03.019> | PubMed
- Tregaskes CA, Harrison M, Sowa AK, van Hateren A, Hunt LG, Vainio O, Kaufman J (2016) Surface expression, peptide repertoire, and thermostability of chicken class I molecules correlate with peptide transporter specificity. *Proc Natl Acad Sci U S A* **113**:692-7 <https://doi.org/10.1073/pnas.1511859113> | PubMed | PubMed Central
- Trowsdale J, Knight JC (2013) Major histocompatibility complex genomics and human disease. *Annu Rev Genomics Hum Genet* **14**:301-23 <https://doi.org/10.1146/annurev-genom-091212-153455> | PubMed | PubMed Central
- Vonrhein C, Flensburg C, Keller P, Sharff A, Smart O, Paciorek W, Womack T, Bricogne G (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D Biol Crystallogr* **67**:293-302 <https://doi.org/10.1107/S0907444911007773> | PubMed | PubMed Central
- Walker BA, Hunt LG, Sowa AK, Skjødtt K, Göbel TW, Lehner PJ, Kaufman J (2011) The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. *Proc Natl Acad Sci U S A* **108**:8396-401 <https://doi.org/10.1073/pnas.1019496108> | PubMed | PubMed Central
- Wallny HJ, Avila D, Hunt LG, Powell TJ, Riegert P, Salomonsen J, Skjødtt K, Vainio O, Vilbois F, Wiles MV, et al. (2006) Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. *Proc Natl Acad Sci U S A* **103**:1434-9 <https://doi.org/10.1073/pnas.0507386103> | PubMed | PubMed Central
- Wei X, Wang S, Li Z, Li Z, Qu Z, Wang S, Zou B, Liang R, Xia C, Zhang N (2021) Peptidomes and Structures Illustrate Two Distinguishing Mechanisms of Alternating the Peptide Plasticity Caused by Swine MHC Class I Micropolymorphism. *Front Immunol* **12** <https://doi.org/10.3389/fimmu.2021.592447> | PubMed | PubMed Central
- Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**:235-42 <https://doi.org/10.1107/S0907444910045749> | PubMed | PubMed Central
- Winter G (2010) xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**:186-190 <https://doi.org/10.1107/S0021889809045701>

Xiao J, Xiang W, Zhang Y, Peng W, Zhao M, Niu L, Chai Y, Qi J, Wang F, Qi P, *et al.* (2018) An Invariant Arginine in Common with MHC Class II Allows Extension at the C-Terminal End of Peptides Bound to Chicken MHC Class I. *J Immunol* **201**:3084-3095 <https://doi.org/10.4049/jimmunol.1800611> | PubMed

Yao Y, Zhao Y, Xu H, Smith LP, Lawrie CH, Watson M, Nair V (2008) MicroRNA profile of Marek's disease virus-transformed T-cell line MSB-1: predominance of virus- encoded microRNAs. *J Virol* **82**:4007-15 <https://doi.org/10.1128/JVI.02659-07> | PubMed | PubMed Central

Zhang J, Chen Y, Qi J, Gao F, Liu Y, Liu J, Zhou X, Kaufman J, Xia C, Gao GF (2012) Narrow groove and restricted anchors of MHC class I molecule BF2*0401 plus peptide transporter restriction can explain disease susceptibility of B4 chickens. *J Immunol* **189**:4478-87 <https://doi.org/10.4049/jimmunol.1200885> | PubMed | PubMed Central

Peer reviews

Reviewer #1 (Public review):

[Editors' note: this version has been assessed by the Reviewing Editor without further input from the original reviewers. The authors have addressed the comments raised in the previous round of review.]

Summary:

Combining in vitro refolding, SEC-based assembly assays, peptide-library screening, MALDI-TOF, LC-MS/MS, structural analysis and immunopeptidomics, this manuscript investigates the peptide-binding principles of the promiscuous chicken MHC-I molecule BF2*21:01.

Strengths:

Although the peptide motif of BF2*21:01 is highly complex, this manuscript identified several principles, including a preference for 10-mer peptides, co-variation between P2 and Pc-2, effects of P3 and Pc-3, and a strong cellular preference for Leu at Pc. The results are important for avian MHC biology and poultry vaccine epitope prediction.

<https://doi.org/10.7554/eLife.111098.2.sa2>

Reviewer #2 (Public review):

Summary:

The study presents an in-depth analysis of the peptide repertoire bound by a promiscuous chicken MHC molecule using mass spectrometry, x-ray crystallography and modelling. While the MHC can bind a very diverse set of peptides, the authors have found some new rules that govern peptide binding to this MHC that could help to build a predictive model to study the repertoire of pathogen-derived peptides.

Strengths:

The study uses a range of well performed experiment across multiple techniques and provides an in-depth analysis of the peptide repertoire, including peptide sequences, length, preferred residues, stability and MHC presentation.

<https://doi.org/10.7554/eLife.111098.2.sa1>

Author response:

The following is the authors' response to the original reviews.

eLife Assessment

This important study investigates the peptide-binding principles of promiscuous chicken MHC molecules. The data from crystallography, mass spectrometry, and modeling are convincing. However, the presentation would benefit from streamlining and clear links between data and conclusions. This paper will be of broad interest to immunologists and those interested in vaccine development.

Overall, we are delighted and grateful to the eLIFE editors and the two reviewers for the careful and thoughtful assessments and reviews of our paper. We are glad that the strengths of the paper were apparent and appreciated. And of course, every paper has weaknesses, especially for a story as complex as this one.

We made only minor changes to accommodate the reviewer comments, along with additions for which we only became aware upon this submission of a revised manuscript. In particular, we shortened the title and abstract to fit what is usual for an eLIFE paper, added Key Resources table with accompanying references, changed the numbering of the figures throughout the manuscript to ensure that each page represented a figure (rather than panels of a figure), moved the figure legends from the embedded figures to a list near the end of the manuscript, and split the supplemental spreadsheet into two renamed Data Source files.

Before answering the comments and questions directly, perhaps a few points would help clarify why the paper is as it is.

First, the experiments cover over three decades of work, with the first gas phase sequencing results done in 1992. Unlike some of the chicken class I alleles which immediately gave completely clear stringent motifs (B4, B12 and B15 in Wallny et al 2006 PNAS, B19 in Han et al 2023 J Immunol), we harvested nothing but confusion from the B21 class I results (Fig. 1). Initially, we thought that the lack of a clear motif for B21 was due to multiple well-expressed class I molecules but only one dominantly-expressed class I molecule was found (Wallny et al 2006 PNAS, Shaw et al 2007 J Immunol) and, to our surprise, bacterially-expressed BF2*21:01 heavy chain and b₂-microglobulin refolded with two synthetic peptides without sequence in common, and the crystal structures showed that this molecule remodeled the binding site to accommodate two such disparate peptides (Koch et al 2008 Immunity). This was the beginning of our understanding of the spectrum of class I alleles from promiscuous generalists to fastidious specialists, which we have explored in a series of further papers (in particular, Chappell et al 2015 eLIFE, Tregaskes et al 2016 PNAS, Kaufman 2018 Trends Immunol, Tregaskes and Kaufman 2022 Mol Immunol).

Second, over these many years, we continued to explore the binding properties of BF2*21:01 in ever more detail, resulting in the current manuscript. We learned only slowly how to probe this unexpected promiscuity, unprecedented in the MHC literature, so that the experiments proceeded with our best understanding at the time, including taking advantage of new approaches as they become available. Each experiment built on the previous set of experiments and each brought us closer to an understanding.

Third, having amassed a collection of data, we chose eLIFE exactly because it allows us to present the entire story from beginning to end without compromise, not just the highlights with the major points illustrated by a few main figures and with the supporting data in many supplementary figures. We include all the data, because it is all part of the story, and so interested researchers to look at the data from their own perspective. Although mostly we provide bar graphs, we include spreadsheets for the raw data (or close to them) for the final experiments (illustrated by Figs. 10 and 14-22) in the two source data files, so these can be assessed easily by others in the field, perhaps using approaches that we may not feel competent to perform.

Public Reviews:

Reviewer #1 (Public review):

Summary:

*Combining in vitro refolding, SEC-based assembly assays, peptide-library screening, MALDI-TOF, LC-MS/MS, structural analysis and immunopeptidomics, this manuscript investigates the peptide-binding principles of the promiscuous chicken MHC-I molecule BF2*21:01.*

Strengths:

*Although the peptide motif of BF2*21:01 is highly complex, this manuscript identified several principles, including a preference for 10-mer peptides, co-variation between P2 and Pc-2, effects of P3 and Pc-3, and a strong cellular preference for Leu at Pc. The results are important for avian MHC biology and poultry vaccine epitope prediction.*

Weaknesses:

The manuscript is sometimes difficult to follow because the authors present a large amount of peptide-library, structural and immunopeptidomics data. without always clearly explaining how these datasets support the proposed simplifying principles.

We are delighted and grateful to the reviewer 1 for the careful and thoughtful comments and questions concerning our manuscript. We are glad that the strengths of the paper were apparent and appreciated, and acknowledge the weaknesses that come with such a complex story with experiments performed over decades.

Major Issues - Points Requiring Clarification or Additional Support:

(1) Line 282-301, 537-545)

*The immunopeptidomics conclusions are mainly based on one B21 cell line with one biological replicate and at least two technical replicates. Given the complexity of the BF2*21:01 peptide repertoire, this is a major limitation. The authors should either provide additional biological replicates or clearly state this limitation in the Abstract, Results and Discussion.*

This limitation is clearly stated in lines 537-545, as part of a paragraph covering the various ways in which the data presented in this manuscript could be improved. In fact, we have performed immunopeptidomics of several different B21 cell types, with many replicates and found similar data as presented, giving us confidence in our interpretations. However, these other experiments belong in different stories, so it is not appropriate that the data be reported in this manuscript.

(2) (Lines 290-313)

*The B21 cell preparations contain both BF2 and the lowly expressed BF1 molecule. Some peptides, especially 8-mers or peptides with atypical motifs, may derive from BF1*21:01. The authors should clarify how BF2*21:01-bound peptides were distinguished from possible BF1-derived peptides, or interpret the immunopeptidomics motif more cautiously. The authors should also provide or cite evidence confirming the B21 haplotype identity of the cell line and chicken materials used for immunopeptidomics.*

The concern about the contribution of BF1*21:01 to the immunopeptidomics is clearly stated in the manuscript, both lines 290-313 and as part of the paragraph describing the limitations of the experiments (lines 542-543). In fact, the expression of BF1 molecules has long been

known to be less than 10% of BF2 molecules at the RNA level, and much less at the protein level (Wallny et al 2006 PNAS, Shaw et al 2007 J Immunol). The proportion of 8mers identified by immunopeptidomics is also low (Fig. 14), and it is not impossible that most 8mers are due to BF1*21:01. We have used assembly assays with peptide libraries, immunopeptidomics and a crystal structure to determine the peptide motif for typical BF1 molecules, of which BF1*21:01 is one and found it may contribute to 8mer peptides but very seldom to longer peptides. This work is unpublished but gives us confidence that the characteristics of BF2*21:01 are not misrepresented by the data in this manuscript.

The sources of the chicken samples and the cell lines are described in detail under Materials and Methods (lines 577-590), citing relevant publications.

(3) (Lines 217-221, 243-253)

The authors acknowledge that MALDI-TOF cannot reliably distinguish peptide combinations with identical or similar masses, nor determine residue positions in some cases. Therefore, MALDI-TOF results should not be overinterpreted as precise evidence for residue preference. The authors should clearly indicate which conclusions are supported by LC-MS/MS.

As described, the experiments follow each other in temporal sequence, so that we started with single peptides, then peptide libraries that varied in one position, then peptide libraries that varied in two positions first analysed by MALDI-TOF and later by LC-MS/MS. The final experiment (Fig. 10, with the original data in the supplementary spreadsheet) directly compares MALDI-TOF and LC-MS/MS results for six peptide libraries, so that the strength of the evidence for residue preference is clear. Throughout the manuscript, we do our best to not to overstate conclusions based on the data of any particular experiment.

(4) (Lines 297-301, 316-330)

The authors suggest that longer peptides may bulge in the middle or extend out of the groove at the C-terminal end. The rationale for the C-terminal extension is not clearly explained. Why is the C-terminal extension considered rather than the N-terminal extension? If the binding register is uncertain, long peptides should be analyzed separately from canonical-length peptides.

When the first sequence of a chicken class I cDNA was determined, an immediate mystery was why one of the so-called invariant residues that coordinate the N- and C-termini of the bound peptide is not conserved (Kaufman et al 1992 J Immunol). In fact, this residue Tyr at position 86 in HLA-A2 and the equivalent position in all mammalian classical class I molecules is an Arg in the classical class I molecules of all non-mammalian vertebrates and is common with class II molecules (Kaufman et al 1995 Semin Immunol). Similar to class II molecules, this Arg in chicken class I molecules allows the peptide to extend out of the C-terminus, as shown by a crystal structure (Xiao et al 2018 J Immunol). The concern that we might be misidentifying the C-terminal amino acid was the basis for the analysis in Figs. 23 and 24, but in the absence of crystal structures, we are not able to provide a final answer this question. Perhaps relevant is the fact that a chicken class II molecule can bind exactly the same peptide in two conformations, one with a canonical 9mer core and the other with an unexpected 10mer core (Goryanin et al 2026 J Virol).

By contrast, N-terminal extensions are only found for some class I alleles and thus far depend on the substitution of small amino acid sidechains for W166 (Li et al 2011 J Virol for bovine, Ma et al 2020 J Immunol for Xenopus, Wei et al 2022 J Immunol for ovine). Thus far, no chicken BF2 sequences have this substitution, consonant with the many crystal structures, including those for BF2*21:01 (Koch et al 2008 Immunity, Chappell et al 2015 eLIFE, this manuscript). However, in unpublished data, we find that most BF1 sequences have sequence

differences that could allow N-terminal extensions, although we have no crystal structures to support this possibility.

(5) (Lines 406-439)

*In vitro assembly assays show that several hydrophobic residues can be tolerated at Pc, whereas immunopeptidomics shows a strong Leu preference at this position. The authors should clarify whether this Leu preference reflects intrinsic BF2*21:01 binding specificity, TAP-mediated peptide transport, antigen processing, peptide loading, or a cell-line-specific effect. Additional experimental support, such as TAP transport analysis, would strengthen this conclusion.*

The preference for Leu at the final position of the peptide by immunopeptidomics of the B21 cell line is strong but not absolute and is certainly affected at the least by the length of the peptide (Figs. 23 and 24). Unpublished immunopeptidomics results (mentioned above) show that this is not a cell line-specific result. The evidence from assembly assays of various peptides is that several hydrophobic amino acids are tolerated with sufficient stability of BF2*21:01 that they are detected in the assay (Figs. 3, 5, 9 and 10). Thermostability assays (Fig. 6) show that peptides with these same hydrophobic amino acids are stable to at least body temperature of chickens. These experiments show that such stability is peptide-dependent (that is, whether a particular amino acid is tolerated depends on the stability conferred by the rest of the peptide). Finally, peptide translocation assays using B21 cells have been done (Tregaskes et al 2016 PNAS) and show that peptides with several hydrophobic amino acids can be pumped into the lumen of the endoplasmic reticulum. However, the assays are with single synthetic peptides, so the data are not extensive enough to separate the effects of the final amino acid from the rest of the peptide. Certainly, peptides with amino acids other than Leu at the C-terminus can be translocated. So, it is not yet clear at which point the preference for Leu at the C-terminus of the peptide arises.

(6) (Lines 172-178, 243-279, 442-457)

The structural analysis explains some residue combinations, such as Arg at P2 with Glu at Pc-2 or Trp at Pc. However, the structural interpretation is not fully integrated with the large-scale peptide library and immunopeptidomics results. Representative high- and low-frequency combinations should be discussed structurally.

Six crystal structures show that BF2*21:02 remodels the binding to accommodate a variety of anchor residues (Koch et al 2008 Immunity, Chappel et al 2015 eLIFE). These crystal structures are representative of sequences found by the immunopeptidomics from very frequent (H-E at roughly 15% 8-12mers) to moderately frequent (E-L at roughly 6% 8-12mers) to infrequent (N-F, A-D and E-D at roughly 1.5%, 1.6% and 0.7% 8-12mers) based on Fig. 18. All but one of the structures has Leu at the C-terminus, with the last one having Val which is found but not frequently by immunopeptidomics.

Similar numbers are found by LC-MS/MS of double-substitution libraries of the two original peptide sequences in Fig. 10 with H-E found frequently (8.1% in P390, 3.8% in P498) and the others infrequently (0.1, 0.9, 1.0, 0.3% in P390, 0, 1.4, 1.0, 0.3% in P498), as calculated from the numbers in the Supplementary data spreadsheet. As discussed in the manuscript, for single-substitution peptide libraries of the two original peptides, Ile/Leu at the C-terminus was very frequent but at the same or slightly less level as Phe, with Met less frequent and Val even less so (Fig. 7).

In addition, there are two more structures along with models explicitly testing some substitutions (Fig. 5). Attempting more current modelling approaches, we found AlphaFold 3 was unable to correctly predict most of the conformations that are found in the crystal structures of BF2*21:01, so we don't feel confident in using them to predict unknown structures of this kind.

(7) The inference of co-variation between P2 and Pc-2, as well as the modulatory effects of P3 and Pc-3, should be better explained. At present, some conclusions appear to be based mainly on residue-frequency patterns, and the logical connection between these observations and the proposed binding principles is not always clear. Statistical analyses, such as mutual information, chi-square tests or permutation tests, and representative structural explanations would strengthen this conclusion.

We endeavored to do our best to explain the data, our interpretations and our reasoning, so we apologise if we have not managed to be as clear as might be desired. We have included as close to raw data as possible for the LC-MS/MS and MALDI-TOF (Fig. 10) and for the immunopeptidomics (Fig. 14 and 18) in the Supplementary Data spreadsheet, exactly so that competent practitioners can carry out further analyses (including the sophisticated statistical tests mentioned).

Reviewer #2 (Public review):

Summary:

The study presents an in-depth analysis of the peptide repertoire bound by a promiscuous chicken MHC molecule using mass spectrometry, x-ray crystallography and modelling. While the MHC can bind a very diverse set of peptides, the authors have found some new rules that govern peptide binding to this MHC that could help to build a predictive model to study the repertoire of pathogen-derived peptides.

Strengths:

The study uses a range of well performed experiment across multiple techniques and provides an in-depth analysis of the peptide repertoire, including peptide sequences, length, preferred residues, stability and MHC presentation.

Weaknesses:

The data overall support the analysis and conclusion well. The only caveat is linked to Figure 4, which does not describe the stability of the peptide-MHC complex, but instead shows refold yield, and the two are not always linked.

We are grateful for the clear understanding of the strengths of the work. With regards to Fig. 4, we agree with the reviewer that there are differences in refold yield but that measure may not be correlated with stability of the peptide-MHC complex. However, we were basing our interpretation of stability on the position and quality of the monomer peak, as illustrated by the trace in Fig. 2, in which a sharp peak at the monomer position represents a stable complex (as seen for the 10 and 11mer peptides) and later peaks represent unstable complexes falling apart during the chromatography (as seen for the 7, 8 and 9mer peptides).

Recommendations for the authors:

Reviewer #1 (Recommendations for the authors):

Minor Issues: Editorial and Data Presentation Modifications

(1) Lines 53-62, 155-170, 303-314: The terms P_C , P_{C-2} and P_{C-3} should be clearly defined early in the manuscript and figure legends.

The Abstract introduces the abbreviations as “peptide positions P_2 and P_{C-2} ” followed by P_C , which are standard usage and seem clear. The first usage in the text is “anchor residues at three positions, with co-variation between the anchor residues at P_2 and P_{C-2} ” which again seems clear, particularly in the context of the text and the figure. However, a parenthetical description has been added to read “...anchor residues at three positions, with co-variation between the anchor residues at P_2 and P_{C-2} (position 2 and the position two before the C-terminus) ...”. Given these usages, it seems unlikely that the reader will fail to understand P_C and P_{C-3} .

(2) Lines 255-279: The term “peptide backbone” should be defined as the fixed sequence context outside the randomized positions, if this is what the authors mean. Suggest clarifying the meaning of “peptide backbone”.

The phrase reads “double-substitution libraries based on four peptide backbones”, which in context of the figures seems clear. However, a parenthetical description has been added: “The examination of double-substitution libraries analysed by MALDI-TOF was expanded to other backbones (that is, other sequences in which two positions were randomised): ...”.

(3) Several figures are complex. The authors should add brief take-home messages to figure legends.

Every figure legend starts with a (sometimes quite long) take-home message. It is not clear what more should be added.

(4) A concise summary table of the proposed binding rules, including preferred peptide length, P_2 , P_{C-2} and P_C preferences, and the effects of P_3/P_{C-3} , would be useful for readers.

A concise summary table of binding rules would be very helpful, but the rules are complex, both qualitative and quantitative. For example, immunopeptidomics shows that 10mers are preferred, but that fails to capture the quantitation. The co-variation of P_2 and P_{C-2} , which is the strongest and best characterized correlation, nevertheless is complex since the occupancy of P_2 by different amino acids (presumably independently of P_{C-2}) varies considerably. At our current level of understanding, it is hard to imagine a table that is both concise and precise. With time and more data, perhaps code for quantitative prediction might be constructed (dare one suggest machine learning...), which is the hope of presenting all the available data in one place.

(5) The Results section contains a large amount of peptide-library, structural and immunopeptidomics data. The authors should improve the logical flow and add clearer transition sentences to explain how each dataset supports the proposed simplifying principles.

The reviewer is of course correct that any written text can be improved (although each critic may have a different idea about which part should be fixed), but we have done our best with the material and time available. We could respond productively to a more detailed critique.

(6) Some statements in the Abstract and Discussion should be softened, especially those related to *in vivo* peptide preferences, BF2*21:01 promiscuity and peptide prediction, given the limited biological replication and uncertainty in peptide assignment.

The statements in the both the Abstract and Discussion are very general, reflecting what we believe to be careful interpretations based on the data. We could respond productively to concerns about specific claims.

Reviewer #2 (Recommendations for the authors):

Overall, the data presented in this study are interesting; however, it is complex, and some results could be merged and simplified, as well as the figures. The data provide an in-depth analysis, using mass spectrometry, of the interplay between the different positions of the peptide and the residues favourable to bind within the antigen-binding cleft.

(1) From the abstract, the concept of "promiscuous generalists and fastidious specialists" is not explored after or defined within the results.

The Abstract introduces the concept of promiscuous generalists and fastidious specialists to provide the basis for exploring the peptide-binding specificity of the most promiscuous class I known, BF2*21:01. This overall concept is described in enormous detail in several publications cited in the current manuscript, but it is not particularly germane to the analyses.

(2) From the abstract "These simplifying principles may eventually allow predictions of pathogen peptides", I'm not sure how "simplifying" the principles are with the data, if anything, it does show a rather complex interplay between the different residues of the peptide that enable the MHC to bind a large and diverse number of peptides.

Compared to any combination of anchor residues being permitted at equal frequency, there are clear preferences which the experiments identify. Instead of an enormous range of possible peptide lengths, roughly 50% of peptides are 10mers. Of course, the structural reasons behind these results are certainly complex and likely must be understood in detail in order to attempt peptide predictions.

(3) Line 48. "less well-expressed". Less than what? Do you mean the level of expression was lower? And if yes, are there values of fold change for comparison?

The sentence in the Abstract reads "Chicken BF2 alleles ... are less well-expressed on the cell surface ... while certain human HLA-B alleles ... are well-expressed ...". Read as a complete sentence, the meaning is clear. This difference for chicken BF2 alleles has been quantified as reported in several publications cited in the current manuscript, ranging from 3-5 fold for peripheral blood lymphocytes to ten-fold for erythrocytes (Kaufman et al 1995 Immunol Rev, Chappel et al 2015 eLIFE), with similar numbers for a few HLA-B alleles on human peripheral blood lymphocytes and monocytes (Chappell et al 2015 eLIFE).

(4) Line 149. "with individual peptides" which peptides are we referring to here?

This introductory sentence to a paragraph outlines the general method used for the experiments in this section of the Results, "refolding *in vitro* ... with individual peptides." Which "individual peptides" are described in the following paragraphs, with the next section of the Results using "refolding *in vitro* ... with peptide libraries".

(5) Line 170. If 9 mers and below are not stable, which is not really quantified or shown with the data on Figure 4, why is refolding material observed for peptides with different lengths of 9 aa and below on Figure 4? A lower yield of refolded material can have a different origin, and there is no association between stability and refold yield. The notion of stability here probably needs to be changed, as it does not apply to the data.

As described in our response to a similar concern above, we are not basing our interpretation of stability on the quantity of refolded material, but on the position and quality of the monomer peak, as described clearly in the legend to Fig. 4: “The original 10mer (REVDEQLLSV) and 11mer (GHAEYGAETL) peptides refold with BF2*21:01 to give stable monomers as do 11mer and 10mer derivative peptides, but 9mer, 8mer or 7mer peptides give heavy chain only peaks” and “The peptides 11mer GHAEYGAETL (top panel), 10mer REVDEQLLSV (middle panel), 11mer GHAEAAAETL and 10mer GHAEAAAETL (bottom panel) gave sharp monomer peaks, while the 9mer GHAEAAETL, 8mer GHAEAAETL and 7mer GHAEETL gave a delayed broad peak indicative of unstable binding or heavy chain.” This concept is illustrated by the trace in Fig. 2 (discussed in the text at the beginning of this section of the Results), in which a sharp peak at the monomer position represents a stable complex, and later peaks represent unstable complexes falling apart during the chromatography or free heavy chains.

| (6) Line 175. As the 3BEV structure had a P2-His, is a comparable structure expected?

This sentence reads “Structures with amino acid substitutions in the 11mer peptide GHAEYGAETL bind with Asp at P_{c-2} and either His or Arg at P₂ (5AD0 and 5ADZ), comparable to the original structure (3BEV) (Fig. 5A).” Minor changes in positions and orientations of individual amino acid sidechains are expected and are clear from the crystal structures presented in Fig. 5A, but are overall comparable to the original peptide in 3BEV, which has a His at P₂ and a Glu at P_{c-2}.

| (7) Line 175 “modelling the substituted”. How was the modelling done?

The sentence reads “Modelling the substituted peptide with Arg at P₂ and Glu at P_{c-2} shows a steric clash that can explain why this peptide did not refold with BF2*21:01 (Fig. 5A).” The legend to Fig. 5A states that “modelling done as detailed in Materials and Methods”, but apparently that section was omitted. A section has been added now to the Materials and Methods which states “Beginning with known crystal structures, modelling was carried out using PyMol with the protein mutagenesis Wizard, the rotamer toggle, show bumps and show surfaces.”

| (8) Line 176 “shows a steric clash” with what? Figure 5 is too small to see, and there is no label on the residue to follow where the steric clash is coming from.

As stated in the legend to Fig. 5, “Structures were determined for GHAEYGAETL (3BEV), GHAEYGAETL (5AD0) and GHAEYGAETL (5ACZ), which all refolded successfully to give stable monomers, while GHAEYGAETL did not (see Fig. 3), all models of which showed steric clashes (one depicted, red arrow).” In the model shown, the clash is between R9 of the BF2*21:01 with the Glu at peptide position 9. Parenthetically, this depiction of the key MHC residues for the co-variation has been used repeatedly, starting with Chappell et al 2015 eLIFE. The picture can be zoomed to make it large enough to see.

| (9) Line 247 “many combinations”. It is not clear here if combinations are referring to a set of double-substitutions or different peptides?

Each peptide in the library has a different double-substitution, so the meaning is the same either way. The point of Fig. 10 is to compare the identification of peptides from LC-MS/MS (in which each peptide is identified exactly) with the identification of sets of peptides from MALDI-TOF (in which the order of the double substitution is not clear, as well as the exact amino acid in the cases of I/L and Q/K). As the figure shows, the numbers are generally very similar, but this sentence describes the percentage of cases for which only one method or the other identified a peptide.

(10) Lines 252-253. *The conclusion of the MS data that the LC-MS/MS is more accurate and sensitive than MALDI-TOF is not very surprising. What was the rationale for using both?*

Our examination of the binding properties of BF2*21:01 for peptides and peptide libraries developed over a long time-span, so in the beginning we looked at single peptides with size exclusion chromatography peaks as the measure, then single- and later double-substitution libraries, first by MALDI-TOF and later by LC-MS/MS. Each set of experiments built on the previous work, so that together they tell the story. For example, after we optimized the use of double-substitution libraries, we were worried about the effects of temperature, so we tested that by MALDI-TOF. While repeating the temperature experiment once we optimized the LC-MS/MS approach might have yielded some additional data, there were other questions to answer.

(11) Line 272-273 *"support the idea that P3 is an important position within the peptide (Figures 8-10) despite not contacting the MHC molecule" The structure of 3BEV clearly shows interaction between the P3-Ala and the Tyr156 of the MHC. Residues that are fully or partially buried in the MHC cleft almost all contact the MHC molecule. Maybe I've missed something, but I think this statement is inaccurate.*

We agree with the reviewer that nearly every peptide residue contacts the MHC molecule (but of course some much more than others). The statement is now changed in the text to read "support the idea that P₃ is an important position within the peptide (Figures 8-10) despite not being an anchor residue."

(12) Line 294. *Figure 14 clearly shows that the number of 8 and 9-mer peptides eluted is at the same level as the 11mer and above, so how does the data fit with the statement that 9mer and shorter peptides are not stable with the MHC?*

Fig. 4 shows that 7, 8 and 9mer derivatives of the original 11mer failed to refold to give a single peak of stable monomers, while Fig. 6 shows that the 10mer derivative of the original 11mer was more thermostable than both the 9mer derivative and the original 11mer. The reason why the 9mer yielded so little monomer in Fig. 4 while giving enough to test by thermostability in Fig. 6 is no longer remembered. A key point is that these results are peptide sequence-specific, so it is not impossible to imagine stable binding of an appropriate 8mer (or perhaps even a 7mer). Another uncertainty, mentioned in the Results and Discussion, is that BF1*21:01 molecules bind primarily 8mers, and the contribution of peptides from BF1*21:01 is not known with certainty.

(13) Line 316. *What was the rationale for choosing peptides > 12aa to see if there is some overhang or bulge? As even 11-12mer can exhibit such features.*

We were just looking for any obvious patterns, but we didn't find any.

(14) *The section starting at line 366 would have benefited from some structure prediction or modelling to illustrate the findings.*

We would have been delighted to model peptides, but we have used AlphaFold3 to compare the models to our crystal structures for seven chicken class I alleles (including BF2*21:01) with one or more peptides. The models sometimes fit the experimental data but they often didn't, often by a wide margin, and with BF2*21:01 the worst (presumably because the system is not trained on MHC molecules which utilise charge transfer). Therefore, we do not feel confident in using any such modelling approaches except in the most defined situations (such as illustrated in Fig. 5).

(15) *Typo - Alleles should be italic, and in vitro as well.*

Alleles of genes are in italics, but alleles of proteins are not. To write *in vitro* in italics is customary, which we have corrected.

(16) Figures

(a) *Some of the figures could be merged together.*

Of course, any presentation can be improved, but which figures we should merge (some already being three pages in length) is not clear. We could productively respond to more detailed suggestions.

(b) *Figure 1. I can't see the different colours mentioned in the figure legend*

Our apologies if the colours are not clear enough, but they are present only as an aid (as elsewhere in the manuscript), with red D and E, blue H, K and R, green N, Q, S and T, and all other amino acids black.

(c) *Figure 12. Is this figure only with 11mer peptides?*

Figure 12 shows the percentage of peptides with particular amino acids at P_2 and P_{C-2} for three 11-mer double-substitution peptide libraries, the sequences of which are written on the graphs and described in the figure legends.

(d) *Table 1. The name of the protein should be added to the table.*

OK.

<https://doi.org/10.7554/eLife.111098.2.sa0>