Clinical & Translational Immunology

Australian and New Zealand Society for Immunology Inc.

Clinical & Translational Immunology 2024; **13**: e70017. doi: 10.1002/cti2.70017
www.wileyonlinelibrary.com/journal/cti

SHORT COMMUNICATION

$CD8^+$ T cell epitope conservation in emerging H5N1 viruses suggests global protection

Emma J Grant[1,2,3] & Stephanie Gras[1,2,3]

[1]Infection and Immunity Program, La Trobe Institute for Molecular Science (LIMS), La Trobe University, Bundoora, VIC, Australia
[2]Department of Biochemistry and Chemistry, School of Agriculture, Biomedicine and Environment (SABE), La Trobe University, Bundoora, VIC, Australia
[3]Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC, Australia

Correspondence
Stephanie Gras, Infection and Immunity Program, La Trobe Institute for Molecular Science (LIMS), La Trobe University, Bundoora, VIC 3086, Australia.
E-mail: s.gras@latrobe.edu.au

Received 17 September 2024;
Revised 5 and 12 November 2024;
Accepted 12 November 2024

doi: 10.1002/cti2.70017

Clinical & Translational Immunology 2024; **13**: e70017

Abstract

Objectives. The recent H5N1 avian influenza outbreak in the USA has sparked fresh fears of avian viruses causing the next pandemic. To date, the H5N1 (clade 2.3.4.4b) outbreak in cattle has spread across several states in the USA, with several humans infected following exposure to cows. This H5N1 clade is also reportedly circulating across Europe, Africa and South America. H5N1 was also detected in a child returning to Australia following travel in India where H5N1 (clade 2.3.2.1a) is also reported to be circulating. There are no licenced vaccines against H5N1 avian influenza viruses for humans. Current vaccines aim to protect against seasonal H1N1 and H3N2 variants are unlikely to provide much protection against the different H5, or other avian viruses. $CD8^+$ T cells are known to provide protection against influenza infection, enhancing viral control and decreasing disease severity. **Methods.** We recently compiled and published a list of the known immunogenic influenza-derived $CD8^+$ T cell epitopes restricted to the most prevalent 10 HLA-A, -B and -C molecules worldwide. We assessed the conservation of a curated list of these influenza A virus-derived $CD8^+$ T cell epitopes in H5N1 viruses' sequences at the heart of the outbreak. **Results.** We identified that > 64% of the $CD8^+$ T cell epitopes are highly conserved (> 90% sequence identity) in the H5N1 viruses, with 60% (18/30) of the most prevalent HLA-I molecules have at least one immunogenic $CD8^+$ T cell epitope conserved in H5N1 viruses. Together these HLA-I molecules with conserved epitopes have a cumulative total of > 100% global coverage. Epitopes derived from the NP, M1, PB2, NS1 and PB1 proteins displayed the highest level of conservation. **Conclusions.** Together, this analysis highlights that globally there is the potential for T cell cross-recognition against the H5N1 viruses that may provide some protection in humans towards the current avian flu outbreak.

Keywords: conservation, epitopes, H5N1, influenza outbreak, T cells

INTRODUCTION

Avian influenza viruses continually circulate in wild bird populations where they cause seemingly limited disease. However, these avian influenza viruses occasionally cause outbreaks in animals farmed for commercial use. H5N1 is one of the 'high pathogenicity avian influenza' (HPAI) viruses that can cause significant morbidity and mortality in poultry. H5N1 viruses have been monitored for decades, with numerous outbreaks in poultry recorded. As reported by the CDC, 912 human cases of H5N1 infection have been reported since 1997, with 29 cases (as reported on 5 June 2024) since January 2022.[1] Recently, a H5N1 virus (clade 2.3.4.4b) has been circulating in the USA, causing outbreaks in cattle across multiple USA states,[2,3] with several humans becoming infected following exposure to infected cows.[4–6] This H5N1 clade is also circulating across several continents including Europe,[7] Africa and South America.[8,9] A similar H5N1 virus (clade 2.3.2.1a) is circulating in South-East Asia,[5] while another (clade 2.3.2.1c) is circulating in Cambodia.[1] There are no licenced human vaccines against avian influenza viruses; however, several recent studies have proposed novel vaccines covering the H5N1 virus.[10–12] Seasonal influenza virus vaccines are typically quadrivalent (although some are trivalent) and contain four virus strains, two influenza A viruses (H1N1 and H3N2) and two influenza B viruses (Yamagata and Victoria) predicted by the WHO Global Influenza Surveillance and Response System (GISRS) as the strains likely to be circulating in the upcoming influenza virus season.[13] These vaccines typically induce a neutralising antibody response against the H1 and H3 Haemagglutinin proteins and as such, these vaccines may not provide much protect against H5 or other avian viruses (H7, H9).

Decades of research have shown that $CD8^+$ T cells are important in the control and clearance of influenza virus infections. In mice, studies showed that $CD8^+$ T cells in the absence of neutralising antibodies could protect against mortality in lethal challenge studies.[14–16] Furthermore, studies in humans have shown correlations against $CD8^+$ T cell numbers and responses with protection.[17–19] Studies in several viruses have also highlighted the importance of $CD8^+$ T cell protection in viral infections, with particular HLA-I molecules being associated with different disease states. This has been shown in HIV, where HIV controller individuals can control the virus more frequently carry the HLA-B*57:01 molecule.[20] More recently, SARS-CoV-2 research has demonstrated links with HLA-B*07:02 expression and 'protection' against COVID-19 severity,[21] while HLA-B*15:01 was shown to be the first genetic association with asymptomatic disease following SARS-CoV-2 infection.[22]

All-in-all, $CD8^+$ T cell responses are important in the control and clearance of viral infection; thus, it would be expected that $CD8^+$ T cell responses could provide a measure of protection in humans against avian influenza viruses should infection occur. Indeed, studies in China of individuals infected with the H7N9 virus showed that the magnitude of $CD8^+$ T cell populations correlated with recovery from severe disease in humans.[19] Several studies have investigated $CD8^+$ T cell responses towards epitopes found in past avian and pandemic influenza virus strains[23–26]; however, there are few studies thus far that have assessed the potential for $CD8^+$ T cell protection against the H5N1 viruses currently circulating across the USA and worldwide.[27]

We recently compiled and published a set of immunogenic influenza-derived $CD8^+$ T cell epitopes restricted to the most prevalent 10 HLA-A, -B and -C molecules expressed worldwide.[13] This gave us the perfect opportunity to assess the level of conservation of these influenza A virus-derived $CD8^+$ T cell epitopes in the H5N1 viruses from clade 2.3.4.4b. We found that > 64% (89/139) of the $CD8^+$ T cell epitopes were conserved in the majority of H5N1 viruses assessed and these peptides were restricted across 18/30 prevalent HLA-I molecules, with > 100% cumulative global coverage. Unsurprisingly, epitopes derived from the highly variable surface-glycoprotein HA, the target of most licenced influenza virus vaccines, were not well conserved (only an average of 54% sequence identity). Conversely, peptides derived from the NP, M1, PB2, NS1 and PB1 proteins had the highest level of conservation (> 90% sequence identity).

Overall, this is some of the first evidence that humans with the most prevalent HLA-I molecules would have pre-existing T cell immunity that should provide protection towards the current H5N1 outbreak.

RESULTS

Protection against H5N1 viruses in > 100% of the global population mediated by CD8$^+$ T cells recognising conserved epitopes presented by prevalent HLA-I molecules

To assess the conservation of known CD8$^+$ T cell epitopes in the recently circulating H5N1 viruses circulating in the USA, we started with our recently published a set of immunogenic influenza-derived CD8$^+$ T cell epitopes restricted to the most prevalent 10 HLA-A, -B and -C molecules expressed worldwide.[13] We downloaded 239 'consensus' H5N1 sequences from clade 2.3.4.4b from the Andersen Lab GitHub repository,[28] translated their DNA sequences to proteins and assessed the level of conservation in the H5N1 viruses using the IEBD conservancy tool.[29] We then curated this list to include only influenza A virus-derived peptides, reformatted this list by protein, combined any identical peptides presented by multiple HLA-I molecules and removed variants of the same peptide presented by the same HLA-I molecule, resulting in 139 influenza A virus-derived peptides for analysis (Table 1). Of the 139 peptides assessed, 89 were conserved (100% sequence similarity in > 90% of the H5N1 sequences analysed), representing 64% of the peptides restricted to the top HLA-I molecules (Figure 1a). To assess the potential protection of CD8$^+$ T cells against H5N1 viruses across the population, we looked at the conservation of peptides by the HLA molecule. We saw strikingly that the conserved epitopes were spread across all the HLA-I molecules that had reported CD8$^+$ T cell epitopes to align. This included nine distinct HLA-A molecules, eight distinct HLA-B molecules and one HLA-C molecule (Figure 1b, Table 2). The frequency of these most prevalent HLA-I molecules worldwide varies considerably; however, the cumulative total of these HLA-I molecules with epitopes conserved in H5N1 viruses (not accounting for co-expression of these HLA-I molecules) is > 100% of the global population (Table 2).

Internal influenza-derived proteins exhibit conservation of known CD8$^+$ T cell epitopes restricted to the top 10 HLA-A, -B and -C molecules in H5N1 viruses

We then assessed the conservation within these CD8$^+$ T cell epitopes by influenza virus protein, where we saw variability in the number of conserved peptides across the different proteins (Figure 1c). As expected, there was more variability than conservation in epitopes derived from the surface glycoproteins HA and NA, with ~16% (n = 2/12) and ~40% (n = 2/5) conserved epitopes, respectively (Figure 1c). Conversely, there was more conservation in epitopes derived from internal proteins such as NP (30/43 = ~70% conserved epitopes), M1 (19/27 = ~70% conserved epitopes), PB2 (9/12 = 75% conserved epitopes), NS1 (6/8 = ~75% conserved epitopes) and PB1 (16/19 = ~84% conserved epitopes) (Figure 1c). Of the 50 epitopes that were not conserved in the H5N1 viruses, varying levels of mutation were observed, from single to up to 10 amino acid changes (Table 1). We determined the predicted binding of the peptides and their mutations using NetMHC4.1[30] and found that the majority of mutations were predicted to decrease the binding ability of the peptide for their reported HLA-I molecule (Table 1, Figure 1d). We also reported the predicted impact of such mutations on peptide presentation by HLA-I molecules, and CD8$^+$ T cell recognition (Table 1).

Lack of M2-derived epitopes conservation in H5N1

Epitopes derived from the M2 protein displayed the least amount of conservation in the H5N1 viruses, with none of the five epitopes conserved (Figure 1c, Table 1). Of these five epitopes, one could not be aligned with the H5N1 sequence (Table 1), two long epitopes were 50% or less sequence identity with IAV (Table 1), and two had a single point mutation (Table 1). The $M2_{45–54}$ epitope mutation was at position 7 changing from an Ile to Val in the H5N1 viruses, both amino acids are small hydrophobic residues and unlikely to impact either HLA binding or TCR interaction. Conversely, The $M2_{70–78}$ epitope also had only a single amino-acid mutation at the last residue, changing from a positively charged long Lys residue to a shorter uncharged Gln that would not be optimal for the HLA-A*11:01 F binding pocket.[31] This mutation is likely to decrease HLA binding as indicated by the decreased predicted peptide-HLA affinity (Table 1).

Limited conservation of HA- and NA-derived epitopes

Unsurprisingly, epitopes derived from the highly variable surface glycoprotein HA displayed limited

Table 1. Conservation of $CD8^+$ T cell epitopes in H5N1 viruses

Protein	HLA molecule	Influenza virus	Peptide sequence	Predicted binding by netMHC	Impact on HLA binding	Impact on T cell recognition
M2	A*03:01/A*11:01/ A*31:01	IAV	RLFFKCIYRR	1.142 = WB/3.633/ 0.234 = SB		
		H5N1 2.3.4.4b	RLFFKC**V**YRR	1.209 = WB/3.848/ 0.268 = SB	*	*
M2	A*11:01	IAV	KSMREEYRK	0.263 = SB		
		H5N1 2.3.4.4b	**E**SMREEYR**Q**	24.077	***	*
M2	B*44:03	IAV	VETPIRNEWGCRCNGSSD	ND		
		H5N1 2.3.4.4b	**ICR**P**TK**N**G**W**E**C**N**C**SD**SSD	ND	***	***
M2	B*44:03	IAV	MSLLTEVETPIRNEWGCR	ND		
		H5N1 2.3.4.4b	MSLLTEVET**YVLSIVPSG**	ND	***	***
M2	B*44:03	IAV	VETPIRNEW	0.007 = SB		
		No alignment		N/A	N/A	N/A
HA	A*02:01/A*02:06	IAV	GLFGAIAGFI	1.323 = WB/3.753		
		H5N1 2.3.4.4b	GLFGAIAGFI	1.323 = WB/3.753	None	None
HA	A*11:01	IAV	RTLDFHDSNVK	0.133 = SB		
		H5N1 2.3.4.4b	RTLDFHDSNVK	0.133 = SB	None	None
HA	A*24:02	IAV	TYPVLNVTM	0.132 = SB		
		H5N1 2.3.4.4b	T**IMEK**NVT**V**	6.712	**	***
HA	A*02:01	IAV	MTIIFLILM	13.306		
		H5N1 2.3.4.4b	M**EN**I**V**L**LA**	27.548	***	***
HA	A*02:01	IAV	RLYQNPTTYI	0.582 = WB		
		H5N1 2.3.4.4b	**N**LY**K**NP**I**TYI	0.883 = WB	*	**
HA	A*11:01	IAV	GIHHPSNSK	0.164 = SB		
		H5N1 2.3.4.4b	GIHH**SN**N**AE**	34.5	**	***
HA	B*44:03	IAV	LENERTLDFHDSNVKNLY	ND		
		H5N1 2.3.4.4b	**M**ENERTLDFHDSNVKNLY	ND	*	*
HA	B*44:03	IAV	AELLVLLENERTLDFHDS	ND		
		H5N1 2.3.4.4b	AELLVL**M**ENERTLDFHDS	ND	*	**
HA	B*44:03	IAV	LENERTLDF	0.243 = SB		
		H5N1 2.3.4.4b	**M**ENERTLDF	0.099 = SB	*	*
HA	A*02:01	IAV	VLLVSLGAI	3.896		
		H5N1 2.3.4.4b	No alignment	N/A	N/A	N/A
HA	A*11:01	IAV	VTAACSHAGK	0.727 = WB		
		H5N1 2.3.4.4b	No alignment	N/A	N/A	N/A
HA	A*11:01	IAV	GIAPLQLGK	0.042 = SB		
		H5N1 2.3.4.4b	No alignment	N/A	N/A	N/A
NA	A*24:02	IAV	SWPDGAELPF	0.227 = SB		
		H5N1 2.3.4.4b	SWPDGAELPF	0.227 = SB	None	None
NA	A*02:01	IAV	CVNGSCFTV	5.403		
		H5N1 2.3.4.4b	CVNGSCFTV	5.403	None	None
NA	A*02:01	IAV	ISIAIGIISLMLQIGNI	ND		
		H5N1 2.3.4.4b	I**CMV**IGI**V**SLMLQIGNI	ND	*	**
NA	A*02:01	IAV	SLCPIRGWAI	4.157		
		H5N1 2.3.4.4b	SLCPI**S**GWAI	3.208	*	**
NA	A*33:03	IAV	RYGNGVWIGR	1.313 = WB		
		H5N1 2.3.4.4b	**K**YGNGVWIGR	1.435 = WB	*	*
NS2	B*40:01/B*40:02	IAV	QEIRTFSFQL	0.281 = SB/0.361 = SB		
		H5N1 2.3.4.4b	QEIRTFSFQL	0.281 = SB/0.361 = SB	None	None
NS2	A*24:02	IAV	TFMQALHLL	0.096 = SB		
		H5N1 2.3.4.4b	TFMQAL**Q**LL	0.176 = SB	*	**
PA	A*11:01/A*24:02	IAV	LYASPQLEGF	20.2/0.057 = SB		
		H5N1 2.3.4.4b	LYASPQLEGF	20.2/0.057 = SB	None	None

(Continued)

Table 1. Continued.

Protein	HLA molecule	Influenza virus	Peptide sequence	Predicted binding by netMHC	Impact on HLA binding	Impact on T cell recognition
PA	A*24:02	IAV	YYLEKANKI	0.012 = SB		
		H5N1 2.3.4.4b	YYLEKANKI	0.012 = SB	None	None
PA	A*02:01	IAV	YINTALLNA	1.515 = WB		
		H5N1 2.3.4.4b	YINTALLNA	1.515 = WB	None	None
PA	A*02:01/A*02:06	IAV	FMYSDFHFI	0.028 = SB/0.044 = SB		
		H5N1 2.3.4.4b	FMYSDFHFI	0.028 = SB/0.044 = SB	None	None
PA	A*02:01	IAV	PPNFSCIENFRAYVDGF	ND		
		H5N1 2.3.4.4b	PPNFS**SL**ENFRAYVDGF	ND	*	*
PA	A*03:01/A*11:01/ A*31:01	IAV	KFLPDLYDYK	1.383 = WB/1.333 = WB/ 1.068 = WB		
		H5N1 2.3.4.4b	KFLPDLYDY**R**	4.179/4.014/0.216 = SB	*	*
NP	A*24:02	IAV	FYIQMCTEL	0.243 = SB		
		H5N1 2.3.4.4b	FYIQMCTEL	0.243 = SB	None	None
NP	A*02:01	IAV	RLIQNSITI	0.285 = SB		
		H5N1 2.3.4.4b	RLIQNSITI	0.285 = SB	None	None
NP	A*11:01	IAV	KTGGPIYRR	0.077 = SB		
		H5N1 2.3.4.4b	KTGGPIYRR	0.077 = SB	None	None
NP	A*01:01/A*26:01	IAV	HSNLNDATY	0.118 = SB/0.678 = WB		
		H5N1 2.3.4.4b	HSNLNDATY	0.118 = SB/0.678 = WB	None	None
NP	A*02:01	IAV	GMDPRMCSL	0.352 = SB		
		H5N1 2.3.4.4b	GMDPRMCSL	0.352 = SB	None	None
NP	A*11:01	IAV	TMVMELIRMIK	4.622		
		H5N1 2.3.4.4b	TMVMELIRMIK	4.622	None	None
NP	A*02:01	IAV	MVMELIRMI	0.759 = WB		
		H5N1 2.3.4.4b	MVMELIRMI	0.759 = WB	None	None
NP	A*02:01	IAV	LIFLARSAL	6.846		
		H5N1 2.3.4.4b	LIFLARSAL	6.846	None	None
NP	A*03:01	IAV	ILRGSVAHK	0.065 = SB		
		H5N1 2.3.4.4b	ILRGSVAHK	0.065 = SB	None	None
NP	A*01:01	IAV	KSCLPACVY	0.878 = WB		
		H5N1 2.3.4.4b	KSCLPACVY	0.878 = WB	None	None
NP	A*02:01/A*02:06	IAV	CLPACVYGL	3.396/7.198		
		H5N1 2.3.4.4b	CLPACVYGL	3.396/7.198	None	None
NP	A*02:01	IAV	QLSTRGVQI	3.226		
		H5N1 2.3.4.4b	QLSTRGVQI	3.226	None	None
NP	A*01:01/B*08:01	IAV	ELRSRYWAI	22.158/0.027 = SB		
		H5N1 2.3.4.4b	ELRSRYWAI	22.158/0.027 = SB	None	None
NP	A*02:01	IAV	SRYWAIRTR	19.942		
		H5N1 2.3.4.4b	SRYWAIRTR	19.942	None	None
NP	A*11:01/A*31:01	IAV	SVQPTFSVQR	0.121 = SB/0.027 = SB		
		H5N1 2.3.4.4b	SVQPTFSVQR	0.121 = SB/0.027 = SB	None	None
NP	A*11:01/A*31:01	IAV	SVQRNLPFER	0.961 = WB/0.128 = SB		
		H5N1 2.3.4.4b	SVQRNLPFER	0.961 = WB/0.128 = SB	None	None
NP	A*02:01/A*02:06	IAV	FQGRGVFEL	0.340 = SB/0.112 = SB		
		H5N1 2.3.4.4b	FQGRGVFEL	0.340 = SB/0.112 = SB	None	None
NP	B*40:02	IAV	GERQNATEI	0.146 = SB		
		H5N1 2.3.4.4b	GERQNATEI	0.146 = SB	None	None
NP	B*15:01	IAV	WHSNLNDATYQRTRALVR	ND		
		H5N1 2.3.4.4b	WHSNLNDATYQRTRALVR	ND	None	None
NP	B*15:01	IAV	HSNLNDATYQR	25.455		
		H5N1 2.3.4.4b	HSNLNDATYQR	25.455	None	None
NP	B*07:02	IAV	LPRRSGAAGA	0.450 = SB		
		H5N1 2.3.4.4b	LPRRSGAAGA	0.450 = SB	None	None

(Continued)

Table 1. Continued.

Protein	HLA molecule	Influenza virus	Peptide sequence	Predicted binding by netMHC	Impact on HLA binding	Impact on T cell recognition
NP	B*58:01	IAV	RGINDRNFW	0.171 = SB		
		H5N1 2.3.4.4b	RGINDRNFW	0.171 = SB	None	None
NP	B*15:01	IAV	IAYERMCNILKGKFQTAA	ND		
		H5N1 2.3.4.4b	IAYERMCNILKGKFQTAA	ND	None	None
NP	B*08:01	IAV	ILKGKFQTA	0.085 = SB		
		H5N1 2.3.4.4b	ILKGKFQTA	0.085 = SB	None	None
NP	B*40:02	IAV	AEIEDLIFL	0.036 = SB		
		H5N1 2.3.4.4b	AEIEDLIFL	0.036 = SB	None	None
NP	B*44:03	IAV	NENPAHKSQLVW	ND		
		H5N1 2.3.4.4b	NENPAHKSQLVW	ND	None	None
NP	B*44:03	IAV	NENPAHKSQLVWMACHSA	ND		
		H5N1 2.3.4.4b	NENPAHKSQLVWMACHSA	ND	None	None
NP	B*15:01	IAV	LELRSRYWAIRTRSGGNT	ND		
		H5N1 2.3.4.4b	LELRSRYWAIRTRSGGNT	ND	None	None
NP	B*15:01	IAV	NQQRASAGQISIQPTFSV	ND		
		H5N1 2.3.4.4b	NQQRASAGQIS**V**QPTFSV	ND	None	None
NP	B*07:02/B*35:01	IAV	LPFERATIM	0.191 = SB/0.023 = SB		
		H5N1 2.3.4.4b	LPFERATIM	0.191 = SB/0.023 = SB	None	None
NP	A*01:01	IAV	CTELKLSDY	0.098 = SB		
		H5N1 2.3.4.4b	CTELKLSD**H**	3.113	***	*
NP	A*02:01	IAV	KLSDYEGRL	0.509 = WB		
		H5N1 2.3.4.4b	KLSD**H**EGRL	0.613 = WB	*	**
NP	A*24:02	IAV	TFLARSALI	0.847 = WB		
		H5N1 2.3.4.4b	**I**FLARSALI	0.609 = WB	*	*
NP	A*03:01/A*11:01	IAV	RVLSFIKGTK	0.080 = SB/0.249 = SB		
		H5N1 2.3.4.4b	RV**S**SFI**R**GT**R**	0.653 = WB/0.994 = WB	**	*
NP	A*02:01	IAV	AMDSNTLEL	0.073 = SB		
		H5N1 2.3.4.4b	**T**MDS**S**TLEL	0.116 = SB	*	**
NP	A*24:02	IAV	PFERATVMAAF	3.301		
		H5N1 2.3.4.4b	PFERAT**I**MAAF	3.501	*	*
NP	B*07:02	IAV	SPIVPSFDM	0.232 = SB		
		H5N1 2.3.4.4b	**N**PIVPSFDM	0.713 = WB	*	*
NP	B*35:01	IAV	PFEKSTIMAAF	28.643		
		H5N1 2.3.4.4b	PFE**RA**TIMAAF	32.25	*	*
NP	B*15:01	IAV	GRFYIQMCTELKLSDYEG	ND		
		H5N1 2.3.4.4b	GRFYIQMCTELKLSD**H**EG	ND	*	**
NP	B*15:01	IAV	NGRKTRIAYERMCNILKG	ND		
		H5N1 2.3.4.4b	NGR**R**TRIAYERMCNILKG	ND	*	*
NP	B*44:03	IAV	YSLIRPNENPAHKSQLVW	ND		
		H5N1 2.3.4.4b	**F**SLIRPNENPAHKSQLVW	ND	*	*
NP	B*15:01	IAV	TMESSTLELRSRYWAIRT	ND		
		H5N1 2.3.4.4b	TM**D**SSTLELRSRYWAIRT	ND	*	*
NP	B*15:01	IAV	GQISIQPTFS	1.859 = WB		
		H5N1 2.3.4.4b	GQIS**V**QPTFS	1.708 = WB	*	*
M1	A*24:02	IAV	LYRKLKREITF	0.321 = SB		
		H5N1 2.3.4.4b	LYRKLKREITF	0.321 = SB	None	None
M1	A*02:01	IAV	LTKGILGFVFTLTVPSE	ND		
		H5N1 2.3.4.4b	LTKGILGFVFTLTVPSE	ND	None	None
M1	A*02:01	IAV	SGPLKAEIAQRLEDV	ND		
		H5N1 2.3.4.4b	SGPLKAEIAQRLEDV	ND	None	None
M1	A*02:01	IAV	LTKGILGFVFTLTVPSERG	ND		
		H5N1 2.3.4.4b	LTKGILGFVFTLTVPSERG	ND	None	None

(Continued)

Table 1. Continued.

Protein	HLA molecule	Influenza virus	Peptide sequence	Predicted binding by netMHC	Impact on HLA binding	Impact on T cell recognition
M1	A*02:01	IAV	KGILGFVFTLTV	12.23		
		H5N1 2.3.4.4b	KGILGFVFTLTV	12.23	None	None
M1	A*02:01	IAV	ILGFVFTLTV	2.066		
		H5N1 2.3.4.4b	ILGFVFTLTV	2.066	None	None
M1	A*02:01	IAV	ILGFVFTLT	5.096		
		H5N1 2.3.4.4b	ILGFVFTLT	5.096	None	None
M1	A*02:01	IAV	GILGFVFTLT	5.107		
	GILGFVFTL	H5N1 2.3.4.4b	GILGFVFTLT	5.107	None	None
M1	A*02:01/A*02:06/ B*07:02/B*35:01/ C*08:01	IAV	GILGFVFTL	0.114 = SB/0.148 = SB/ 10.849/14.876/3.147		
		H5N1 2.3.4.4b	GILGFVFTL	0.114 = SB/0.148 = SB/ 10.849/14.876/3.147	None	None
M1	A*02:01	IAV	ALASCMGLI	2.759		
		H5N1 2.3.4.4b	ALASCMGLI	2.759	None	None
M1	A*11:01	IAV	SCMGLIYNR	2.336		
		H5N1 2.3.4.4b	SCMGLIYNR	2.336	None	None
M1	A*11:01	IAV	LASCMGLIYNRMG	ND		
		H5N1 2.3.4.4b	LASCMGLIYNRMG	ND	None	None
M1	A*11:01	IAV	GALASCMGLIYNR	ND		
		H5N1 2.3.4.4b	GALASCMGLIYNR	ND	None	None
M1	A*11:01	IAV	ALASCMGLIYNRM	ND		
		H5N1 2.3.4.4b	ALASCMGLIYNRM	ND	None	None
M1	A*03:01/A*11:01/ A*31:01	IAV	ASCMGLIYNR	3.067/0.702 = WB/ 0.199 = SB		
		H5N1 2.3.4.4b	ASCMGLIYNR	3.067/0.702 = WB/ 0.199 = SB	None	None
M1	A*11:01	IAV	RMVLASTTAK	0.197 = SB		
		H5N1 2.3.4.4b	RMVLASTTAK	0.197 = SB	None	None
M1	A*33:03	IAV	LASCMGLIYN	86.667		
		H5N1 2.3.4.4b	LASCMGLIYN	86.667	None	None
M1	B*35:01	IAV	ASCMGLIY	8.243		
		H5N1 2.3.4.4b	ASCMGLIY	8.243	None	None
M1	B*40:01/B*40:02/B*44:03	IAV	TEVETYVLSI	0.685 = WB/0.910 = WB/ 0.962 = WB		
		H5N1 2.3.4.4b	TEVETYVLSI	0.685 = WB/0.910 = WB/ 0.962 = WB	None	None
M1	A*11:01/A*24:02	IAV	AYQKRMGVQM	24.846/1.169 = WB		
		H5N1 2.3.4.4b	AYQKRMGVQ**L**	27.5/0.634 = WB	*	*
M1	A*24:02	IAV	TFHGAKEVSL	3.026		
		H5N1 2.3.4.4b	TFHGAKEV**A**L	3.75	*	*
M1	A*11:01	IAV	SCMGLIYNRMGAV	ND		
		H5N1 2.3.4.4b	SCMGLIYNRMG**T**V	ND	*	*
M1	A*11:01	IAV	ASCMGLIYNRMGA	ND		
		H5N1 2.3.4.4b	ASCMGLIYNRMG**T**	ND	*	*
M1	A*11:01	IAV	SIIPSGPLK	0.006 = SB		
		H5N1 2.3.4.4b	SI**V**PSGPLK	0.013 = SB	*	*
M1	B*35:01	IAV	AGALASCMGLIYNRMGA	ND		
		H5N1 2.3.4.4b	**T**GALASCMGLIYNRMG**T**	ND	*	*
M1	B*40:01/B*40:02	IAV	SEQAAEAMEV	1.434 = WB/1.924 = WB		
		H5N1 2.3.4.4b	SEQA**V**EAMEV	1.484 = WB/1.747 = WB	*	**
M1	A*02:01	IAV	IMDKNIILKA	2.074		
		H5N1 2.3.4.4b	No alignment	N/A	N/A	N/A

(Continued)

Table 1. Continued.

Protein	HLA molecule	Influenza virus	Peptide sequence	Predicted binding by netMHC	Impact on HLA binding	Impact on T cell recognition
PB2	A*11:01	IAV	SSSFSFGGFTFK	16		
		H5N1 2.3.4.4b	SSSFSFGGFTFK	16	None	None
PB2	A*03:01/A*11:01/ A*31:01	IAV	SFSFGGFTFK	0.091 = SB/0.019 = SB/ 0.732 = WB		
		H5N1 2.3.4.4b	SFSFGGFTFK	0.091 = SB/0.019 = SB/ 0.732 = WB	None	None
PB2	A*11:01	IAV	FSFGGFTFK	0.037 = SB		
		H5N1 2.3.4.4b	FSFGGFTFK	0.037 = SB	None	None
PB2	A*24:02	IAV	TYQWIIRNW	0.062 = SB		
		H5N1 2.3.4.4b	TYQWIIRNW	0.062 = SB	None	None
PB2	A*24:02	IAV	QYSGFVRTL	0.053 = SB		
		H5N1 2.3.4.4b	QYSGFVRTL	0.053 = SB	None	None
PB2	A*03:01/A*11:01/ A*31:01	IAV	VLRGFLILGK	0.262 = SB/2.635/2.951		
		H5N1 2.3.4.4b	VLRGFLILGK	0.262 = SB/2.635/2.951	None	None
PB2	A*24:02	IAV	RYGPALSI	0.514 = WB		
		H5N1 2.3.4.4b	RYGPALSI	0.514 = WB	None	None
PB2	B*44:03	IAV	GRQEKNPALRMKWMMAMK	ND		
		H5N1 2.3.4.4b	GRQEKNPALRMKWMMAMK	ND	None	None
PB2	B*44:03	IAV	QEKNPALRMKW	0.053 = SB		
		H5N1 2.3.4.4b	QEKNPALRMKW	0.053 = SB	None	None
PB2	A*24:02	IAV	HYPKIYKTYF	0.032 = SB		
		H5N1 2.3.4.4b	HYPK**V**YKTYF	0.030 = SB	*	*
PB2	A*02:01	IAV	PVAGGTSSIYI	25.22		
		H5N1 2.3.4.4b	PVAGGTSS**V**YI	25.659	*	*
PB2	A*02:01	IAV	SLENFRAYV	0.734 = WB		
		H5N1 2.3.4.4b	No alignment	N/A	N/A	N/A
NS1	A*02:01/A*02:06	IAV	FQVDCFLWHV	0.577 = WB/0.146 = SB		
		H5N1 2.3.4.4b	FQVDCFLWHV	0.577 = WB/0.146 = SB	None	None
NS1	A*02:01	IAV	DQAIMDKNIILKANFSV	ND		
		H5N1 2.3.4.4b	DQAIMDKNIILKANFSV	ND	None	None
NS1	A*02:01	IAV	AIMDKNIIL	0.299 = SB		
		H5N1 2.3.4.4b	AIMDKNIIL	0.299 = SB	None	None
NS1	B*44:03	IAV	WNDNTVRVSETLQRFAWR	ND		
		H5N1 2.3.4.4b	WNDNTVRVSETLQRFAWR	ND	None	None
NS1	B*44:03	IAV	SETLQRFAW	0.012 = SB		
		H5N1 2.3.4.4b	SETLQRFAW	0.012 = SB	None	None
NS1	B*44:03	IAV	QEIRTFSFQL	0.538 = WB		
		H5N1 2.3.4.4b	QEIRTFSFQL	0.538 = WB	None	None
NS1	A*02:01	IAV	IVDKNITLKA	6.706		
		H5N1 2.3.4.4b	I**M**DKNI**I**LKA	2.074	**	*
NS1	B*44:03	IAV	RVSETLQRFAWRSSNENG	ND		
		H5N1 2.3.4.4b	RVSETLQRFAWRSSNE**D**G	ND	*	**
PB1	A*11:01	IAV	AVATTHSWIPK	0.280 = SB		
		H5N1 2.3.4.4b	AVATTHSWIPK	0.280 = SB	None	None
PB1	A*24:02	IAV	RYGFVANF	0.195 = SB		
		H5N1 2.3.4.4b	RYGFVANF	0.195 = SB	None	None
PB1	A*01:01/A*26:01	IAV	YSHGTGTGY	0.119 = SB/0.080 = SB		
		H5N1 2.3.4.4b	YSHGTGTGY	0.119 = SB/0.080 = SB	None	None
PB1	A*02:01	IAV	GMMMGMFNMLSTVLGVS	ND		
		H5N1 2.3.4.4b	GMMMGMFNMLSTVLGVS	ND	None	None
PB1	A*02:01	IAV	FNMLSTVLGV	0.871 = WB		
		H5N1 2.3.4.4b	FNMLSTVLGV	0.871 = WB	None	None

(Continued)

Table 1. Continued.

Protein	HLA molecule	Influenza virus	Peptide sequence	Predicted binding by netMHC	Impact on HLA binding	Impact on T cell recognition
PB1	A*02:01/A*02:06/ A*11:01	IAV	NMLSTVLGV	0.226 = SB/0.495 = SB/23		
		H5N1 2.3.4.4b	NMLSTVLGV	0.226 = SB/0.495 = SB/23	None	None
PB1	A*03:01/A*11:01/ A*31:01	IAV	KLVGINMSKK	0.073 = SB/0.845 = WB/ 3.880		
		H5N1 2.3.4.4b	KLVGINMSKK	0.073 = SB/0.845 = WB/ 3.880	None	None
PB1	A*03:01/A*11:01/ A*31:01	IAV	GTFEFTSFFY	1.202 = WB/0.429 = SB/ 5.033		
		H5N1 2.3.4.4b	GTFEFTSFFY	1.202 = WB/0.429 = SB/ 5.033	None	None
PB1	A*24:02	IAV	FYRYGFVANF	0.399 = SB		
		H5N1 2.3.4.4b	FYRYGFVANF	0.399 = SB	None	None
PB1	A*02:01	IAV	FVANFSMEL	0.140 = SB		
		H5N1 2.3.4.4b	FVANFSMEL	0.140 = SB	None	None
PB1	A*01:01/A*26:01	IAV	LVSDGGPNLY	0.008 = SB/0.415 = SB		
		H5N1 2.3.4.4b	LVSDGGPNLY	0.008 = SB/0.415 = SB	None	None
PB1	A*01:01	IAV	VSDGGPNLY	0.002 = SB		
		H5N1 2.3.4.4b	VSDGGPNLY	0.002 = SB	None	None
PB1	B*07:02	IAV	QPEWFRNVL	0.063 = SB		
		H5N1 2.3.4.4b	QPEWFRNVL	0.063 = SB	None	None
PB1	B*15:01	IAV	KMARLGKGY	0.129 = SB		
		H5N1 2.3.4.4b	KMARLGKGY	0.129 = SB	None	None
PB1	B*07:02	IAV	LPSFGVSGI	0.679 = WB		
		H5N1 2.3.4.4b	LPSFGVSGI	0.679 = WB	None	None
PB1	B*15:01	IAV	TQIQTRRSF	0.019 = SB		
		H5N1 2.3.4.4b	TQIQTRRSF	0.019 = SB	None	None
PB1	A*24:02	IAV	RYTKTTYWW	0.075 = SB		
		H5N1 2.3.4.4b	**K**YTKTTYWW	0.077 = SB	*	*
PB1	A*02:01	IAV	TVIKTNMI	21.605		
		H5N1 2.3.4.4b	TVIK**N**NMI	27	*	**
PB1	A*24:02	IAV	EWMSIRPYF	0.129 = SB		
		H5N1 2.3.4.4b	No alignment	N/A	N/A	N/A

Overall, 239 'consensus' H5N1 sequences from clade 2.3.4.4b were downloaded from the Andersen Lab GitHub repository[28] their DNA sequences were translated into proteins using the Expasy Translate Tool.[42] $CD8^+$ T cell epitopes were obtained from Leong *et al.*[13] and conservation of these epitopes in H5N1 viruses was determined using the IEBD Conservancy Analysis tool.[29] This list was then curated as per the Methods. Mutations are shown in bold. Conservation: epitopes conserved in H5N1 viruses are shown in green, epitopes unique to H5N1 are shown in yellow while epitopes that could not align are shown in orange. NetMHC peptide-binding predictions: Binding predictions (%Rank_EL) and bind level where SB refers to strong binder while WB refers to weak binder. Sequences too long for NetMHC prediction (> 11 amino acids) are denoted with ND as not determined. Impact: no impact is denoted as 'none', weak or unlikely impact is denoted with a *, moderate impact as ** and strong impact as ***, while NA refers to not available because of no alignment with IAV viruses being found.

conservation in H5N1 viruses, with ~16% conserved epitopes (2/12 epitopes) (Figure 1c, Table 1). Of the 10 epitopes that were not conserved in H5N1, three could not be aligned with the H5N1 viruses (Table 1), three contain a single point mutation (Table 1), while the remaining four had between three and five mutations (Table 1). Of the seven mutant epitopes 'unique' to H5N1 (epitopes that were not conserved but where homologous peptides could be identified, denoted in yellow, Table 1), 3 were expected to impact HLA presentation either moderately ($HA_{176–184}$ and $HA_{195–203}$), or strongly ($HA_{1–9}$). Four of the seven unique HA-derived epitopes had at least three mutations, and three of these are likely to strongly impact T cell recognition. $HA_{176–184}$ residues 4 (Val to Glu) and 5 (Leu to Lys) mutations will change the peptide from having small hydrophobic residues to having large and charged residues in the central part of

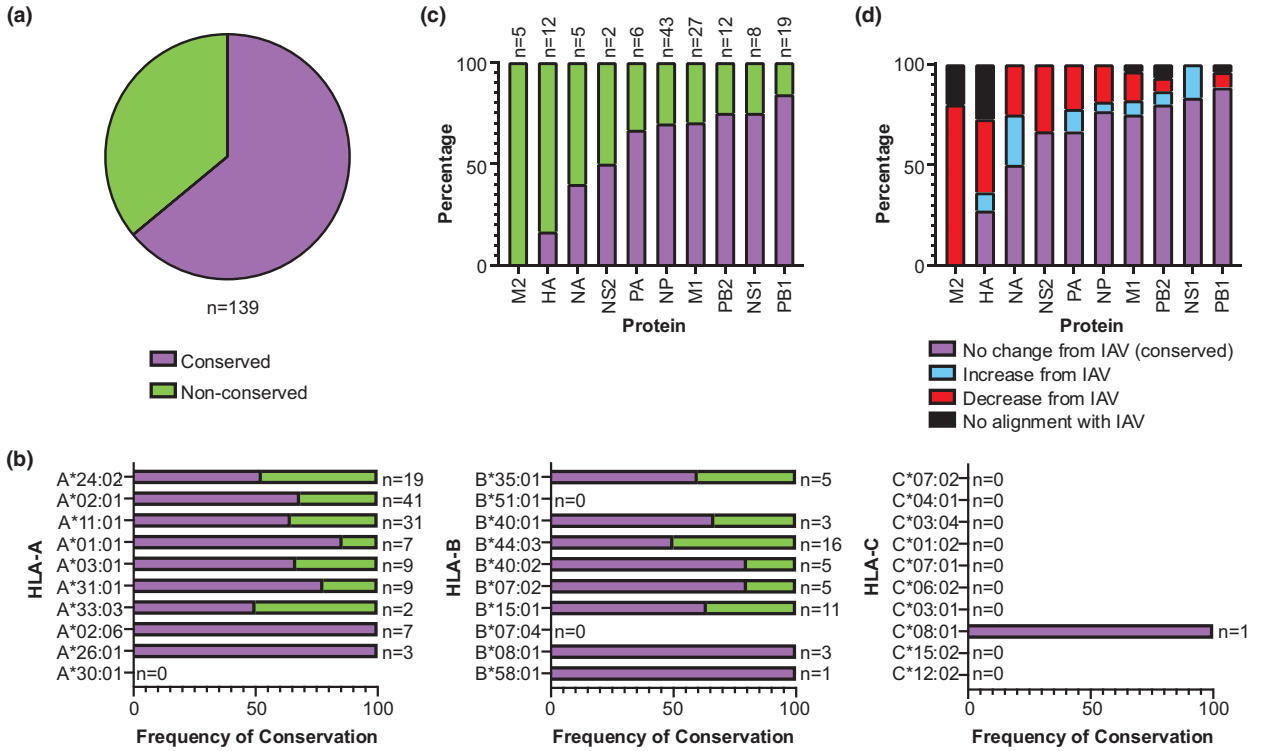

Figure 1. Conservation in CD8$^+$ T cell epitopes restricted to the top 10 HLA-A, -B and -C molecules in H5N1 viruses. Overall, 239 'consensus' H5N1 sequences from clade 2.3.4.4b circulating in the USA were downloaded from the Andersen Lab GitHub repository[28] their DNA sequences were translated into proteins using the Expasy Translate Tool.[42] CD8$^+$ T cell epitopes restricted to the most prevalent 10 HLA-A, -B and -C molecules worldwide were obtained from Leong *et al.*[13] as was curated as per the Methods. Conservation in this curated set of epitopes was determined using the IEBD Conservancy Analysis tool.[29] An epitope was deemed 'conserved' if the same sequence is identified across > 90% of the consensus sequences. Predicted binding (%Rank_EL) and bind level for the relevant HLA molecule was estimated using NetMHC4.1.[30] **(a)** Frequency of conservation of the curated list of CD8$^+$ T cell epitopes in the analysed H5N1 viruses. **(b)** Conservation of CD8$^+$ T cell epitopes by the HLA-I molecule. **(c)** Conservation of CD8$^+$ T cell epitopes by influenza A virus protein. **(d)** Effect of mutations in H5N1-derived epitopes on the binding prediction (where binding prediction could be done or no alignment with IAV viruses was found) for their reported HLA-I, expressed as no change (purple), increase from IAV (red), decrease from IAV (blue) or no alignment with IAV (black).

the peptide. Located centrally on the peptide, these mutations are likely to impact TCR recognition.[32] Similarly, the $HA_{195–203}$ bears mutations at positions 5 (Pro to Ser) and 6 (Ser to Asn) that are likely to strongly impact both the conformation of the peptide and the TCR interaction as well.[32] Finally, the $HA_{1–9}$ epitope had five residues mutated in H5N1, with both primary anchor residues mutated from a small P2-Thr to a large and charge P2-Glu, and from a large P9-Met to a small hydrophobic P9-Ala. Both mutations will have a strong impact on the peptide stability in the cleft of HLA-A*02:01. In addition, central residue in IAV is a large aromatic P5-Phe that is mutated in H5N1 in a small hydrophobic P5-Val, likely to impact on TCR recognition and binding (Table 1).

Likewise, epitopes derived from the surface NA protein displayed only ~40% conserved epitopes (2/5 peptides) (Figure 1c, Table 1). Of the three non-conserved epitopes, two had one mutation and the other had four mutations. None of these mutations are predicted to decrease the peptide affinity for the HLA-I molecule, or HLA presentation, and two may have a moderate impact on T cell recognition.

More than half of the epitopes derived from NS2, PA, NP, M1, PB2 and NS1 proteins are conserved in H5N1

There are only two NS2-derived epitopes reported with one conserved in H5N1 viruses and the other had one mutation at position 7 (His to Gln) not

Table 2. Conservation in influenza A virus-derived $CD8^+$ T cell epitopes by top HLA-I molecules

	HLA-A			HLA-B			HLA-C		
HLA rank	Allomorph	No. of conserved epitopes/total epitopes	HLA-I Frequency %	Allomorph	No. of conserved epitopes/total epitopes	HLA-I frequency %	Allomorph	No. of conserved epitopes/total epitopes	HLA-I frequency %
1	A*24:02	10/19	18.82	B*35:01	3/5	5.47	C*07:02	0/0	13.10
2	A*02:01	28/41	15.28	B*51:01	0/0	5.22	C*04:01	0/0	11.18
3	A*11:01	20/31	11.66	B*40:01	2/3	5.12	C*03:04	0/0	9.13
4	A*01:01	6/7	4.84	B*44:03	8/16	4.47	C*01:02	0/0	8.48
5	A*03:01	6/9	4.27	B*40:02	4/5	4.18	C*07:01	0/0	6.89
6	A*31:01	7/9	4.09	B*07:02	4/5	4.11	C*06:02	0/0	6.16
7	A*33:03	1/2	4.08	B*15:01	7/11	3.43	C*03:03	0/0	5.58
8	A*02:06	7/7	3.47	B*07:04	0/0	3.12	C*08:01	1/1	4.52
9	A*26:01	3/3	3.35	B*08:01	3/3	2.96	C*15:02	0/0	3.36
10	A*30:01	0/0	2.51	B*58:01	1/1	2.89	C*12:02	0/0	3.19
Cumulative global frequency of HLA-I with conserved epitopes			69.86			32.63			4.52

Overall, 239 'consensus' H5N1 sequences from clade 2.3.4.4b circulating in the USA were downloaded from the Andersen Lab GitHub repository[28] their DNA sequences were translated into proteins using the Expasy Translate Tool.[42] $CD8^+$ T cell epitopes restricted to the most prevalent 10 HLA-A, -B and -C molecules worldwide were obtained from Leong *et al.*[13] and conservation of these epitopes in H5N1 viruses was determined using the IEBD Conservancy Analysis tool.[29] We then curated this list as per the Methods, resulting in 139 influenza A virus-derived peptides for analysis (Table 1).

predicted to affect peptide affinity or presentation by HLA (Figure 1c, Table 1), and that might have a moderate impact on TCR binding.

In the PA protein, six epitopes were identified, of which four (~67%) were conserved in H5N1 viruses (Figure 1c, Table 1). The two non-conserved epitopes contained one or two mutations. The $PA_{104–113}$ has one mutation at the last residue, from a Lys to Arg. Both residues are long and positively charged and should not impact on the peptide binding or TCR recognition. The 17mer $PA_{220–236}$ has two mutations in the centre of the peptide at P6 (Cys to Ser) and P7 (Ile to Leu) that would not impact the peptide binding and have a minimal impact on TCR recognition (Table 1).

The NP protein had the largest number of epitopes characterised ($n = 43$) of which ~70% ($n = 30$) were conserved in H5N1 (Figure 1c, Table 1). The remaining 13 had one ($n = 10$), two ($n = 2$) or three ($n = 1$) mutations, with various levels of effect on predicted peptide-HLA-I affinity. Only two epitopes had mutation at one of the main anchor residues, namely $NP_{342–351}$ and $NP_{44–52}$. In $NP_{342–351}$ the P10-Lys mutation to Arg is unlikely to impact peptide binding, as both residues are long and positively charged (Table 1). For $NP_{44–52}$, the mutation might impact the peptide affinity as it changes the P9-Tyr to a P9-His, which is smaller and charged. Only a few mutations might have an impact on the T cell response (Table 1) and, overall, a high level of cross-protective T cell immunity should be observed towards H5N1 NP-derived epitopes.

Twenty-seven epitopes derived from the M1 protein were identified of which 19 were conserved (70% conservation) in H5N1 (Figure 1c, Table 1). Of the eight non-conserved peptides, one could not be aligned with H5N1, while six of the remaining seven had a single mutation and one had two mutations. None of the mutations in these epitopes are expected to have an impact on peptide affinity or presentation, and only the mutation in the $M1_{196–205}$ epitope (P5-Ala to P5-Val) may have a moderate impact on TCR recognition.

Twelve epitopes were identified in the PB2 protein, of which 75% (9/12 epitopes) were conserved in H5N1 (Figure 1c, Table 1). One epitope had no homologous peptides identified. The two mutated epitopes unique to H5N1, had a single amino acid mutation, and, as the mutations were not within an anchor residue, they are not expected to impact peptide binding. In addition, both mutations are from an Ile to a Val, both small hydrophobic residues that are unlikely to impact on TCR recognition.

From the NS1 protein six of eight peptides are conserved in H5N1 viruses (Figure 1c, Table 1).

One mutant, $NS1_{193–210}$ had a single mutation at the second last residue (Asn to Asp) that would have no impact on peptide binding and might have a moderate impact on T cell recognition. The $NS1_{115–124}$ epitope has two mutations in H5N1, one at P2 anchor residue from a small Val to a larger Met that could decrease peptide-binding affinity for the HLA, and the other mutation is at P8-Thr from P8-Ile that are both small residues (Table 1).

The epitopes from PB1 are the most conserved in H5N1

In the PB1 protein, 16 of the 19 (~84%) characterised epitopes were conserved in the H5N1 viruses assessed (Figure 1c, Table 1). One peptide had no homologous peptide identified in H5N1, and two had a single mutation. The $PB1_{430–438}$ has a mutation at P1 from Arg to Lys that should not impact on peptide or TCR binding, while $PB1_{528–535}$ has its P5-Thr mutated to a larger P5-Asn that could impact TCR recognition (Table 1).

Overall, these analyses show that although there was no conservation in ~36% (50/139) of the epitopes, the mutations may not actually prevent HLA-I binding or presentation or a $CD8^+$ T cell response, and thus the potential for pre-existing immunity may be even higher.

DISCUSSION

Avian influenza viruses circulate continually through wild bird populations and occasionally cause localised outbreaks in poultry and more recently cattle.[1] Although rare, these avian influenza viruses can occasionally transmit into humans following close and prolonged contact with infected animals.[1,4] $CD8^+$ T cells are known to provide protection against influenza viruses infection[14,15,17–19] and since no avian influenza virus vaccines exist for humans, we wanted to determine the potential for pre-existing immunity towards the H5N1 clade 2.3.4.4b virus in humans. We determined the conservation of known influenza-derived $CD8^+$ T cell epitopes restricted to highly prevalent HLA-A, -B and -C molecules[13] in H5N1 clade 2.3.4.4b viruses at the centre of the USA outbreak.[1,28] It is important to note that although these epitopes have been scientifically validated as immunogenic in previous studies, for the purposes of this analysis, we are assuming that these epitopes will still be processed and presented in the current clade 2.3.4.4b H5N1 viruses. Surprisingly, 64% of $CD8^+$ T cell epitopes restricted to the top 10 prevalent HLA-A, -B and -C molecules were conserved in > 90% of the H5N1 viruses. Conservation of $CD8^+$ T cell epitopes restricted to the top 10 HLA-A, -B and -C molecules in H5N1 viruses was spread across 18 different HLA-I molecules, with a global cumulative coverage of > 100%, suggesting that most individuals worldwide have some level of protection against these H5N1 viruses because of $CD8^+$ T cell responses and prevalent HLA-I molecules. This is similar to our previous study which estimated that pre-existing immunity towards the avian-derived H7N9 virus occurred in 16–57% of the population depending on ethnicity.[24] Furthermore, in the $CD8^+$ T cell epitopes that were not conserved in the H5N1 viruses assessed (50/139 epitopes) varying levels of mutation were observed, some of which are not expected to impact either HLA presentation or $CD8^+$ T cell recognition, and, as such, the potential for pre-existing immunity may be even higher because of $CD8^+$ T cell cross-reactivity which can occur towards similar variant peptides.[26,33]

There was generally a higher level of conservation in epitopes derived from internal influenza proteins, suggestive of their important role in the virus life cycle. There was > 67% conservation in H5N1 viruses within epitopes derived from PA, NP, M1, PB2, NS1 and PB1. Interestingly, NP is considered one of the most immunogenic proteins from influenza A virus,[34,35] and contains many well characterised and highly immunogenic peptides presented by a range of HLA-I molecules.[33–36] These include the highly immunogenic HLA-A*03:01-restricted $NP_{265–273}$[24,35,36] and HLA-B*08:01-restricted $NP_{225–233}$[24,37] epitopes determined as 'universal' for their conservation in past H7N9 and H5N1 avian viruses.[24] Also included are the highly immunogenic HLA-A*11:01-restricted $NP_{91–99}$ peptide[35,38] and the HLA-B*07:02/-B*35:01-restricted $NP_{418–426}$[25,26,39,40] peptide, which has likewise been seen in past avian and pandemic influenza virus strains. All these highly immunogenic peptides were 100% conserved in the H5N1 viruses assessed in this study. Similarly, the M1 protein is considered highly immunogenic and is home to several highly immunogenic peptides, including the most well studied and highly conserved HLA-A*02:01 restricted 'universal' $M1_{58–66}$ peptide.[24,40,41]

Conversely, there was less conservation in epitopes derived from the surface HA and NA glycoproteins, with the HA being the target of most licenced influenza vaccines worldwide,[13] suggesting that vaccines designed to protect against seasonal influenza viruses are unlikely to provide any protection against H5N1 viruses. Collectively, this suggests that we should consider including more conserved internal influenza-derived proteins in future influenza vaccines to induce strain cross-protective $CD8^+$ T cell responses alongside neutralising antibody responses towards HA, as this may assist in protection against various influenza virus strains including avian-derived influenza viruses.

Overall, this analysis suggests that most of the global population could have a level of T cell cross-reactivity that recognises conserved epitopes that could provide protection against the clade 2.3.4.4b H5N1 viruses assessed, should sporadic infections of humans occur.

METHODS

H5N1 consensus sequences

Overall, 239 'consensus' H5N1 sequences from clade 2.3.4.4b circulating in the USA were downloaded from the Andersen Lab GitHub repository on 7–8 May 2024.[28] Their DNA sequences were translated into proteins using the Expasy Translate Tool.[42] Unknown amino acids denoted as an 'X' following translation using the Expasy Translate Tool were removed from the sequences before conservation analysis.

Influenza-derived epitopes, conservation analysis and peptide affinity

$CD8^+$ T cell epitopes restricted to the most prevalent 10 HLA-A, -B and -C molecules worldwide were obtained from Leong *et al*.[13] Conservation of epitopes against the H5N1 viruses was determined using the IEBD Conservancy Analysis tool.[29] An epitope was deemed 'conserved' if the same sequence is identified across > 90% of the consensus sequences. The epitope list was then curated as follows. Epitopes with an identical amino acid sequence but different HLA-I restriction were considered one epitope. Epitopes that were variants of the same peptide and restricted to the same HLA-I molecule were also considered a single epitope. In this latter case, the representative epitope was selected if it was 100% conserved with the H5N1 sequences, or in the absence of 100% conservation, the epitope with the least amount of amino acid changes was selected as the representative epitope. Epitopes with post-translational modifications were also removed. Epitopes used in the analysis are indicated in Table 1. Epitopes are grouped by protein including PB1 (including Polymerase Basic Protein 1 and RNA Polymerase), PB2 (Polymerase Basic Protein 2), PA (Acid Polymerase), NP (Nucleoprotein), NS1 (Non-Structural Protein 1), NS2 (Non-Structural Protein 2 and Nuclear Export Protein), M1 (Matrix 1), M2 (Matrix 2), HA (Haemagglutinin) and NA (Neuraminidase). Predicted binding of the peptide for the relevant HLA molecule (% Rank_EL and binding score) were determined using NetMHC4.1[30] and are reported in Table 1.

ACKNOWLEDGMENTS

We thank the Andersen lab for making these consensus H5N1 strains available for analysis. We also thank the NIH for their development and maintenance of the online IEDB conservancy analysis tool. We are also grateful for the publicly available online databases EXPASY translate tool and NetMHC4.1. EJG was supported by an Australian Research Council (ARC) DECRA Fellowship (DE210101479). SG was supported by a National Health and Medical Research Council (NHMRC) Senior Research Fellowship (#1159272). Open access publishing facilitated by La Trobe University, as part of the Wiley - La Trobe University agreement via the Council of Australian University Librarians.

AUTHOR CONTRIBUTIONS

Emma J Grant: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; validation; visualization; writing – original draft; writing – review and editing. **Stephanie Gras:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; supervision; validation; visualization; writing – original draft; writing – review and editing.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study and an Excel spreadsheet of Table 1 are available from the corresponding author upon reasonable request.

REFERENCES

1. Technical Report: June 2024 Highly Pathogenic Avian Influenza A (H5N1) Viruses. Available from: https://www.cdc.gov/bird-flu/php/technical-report/h5n1-06052024.html
2. Sah R, Srivastava S, Kumar S *et al*. Concerns on H5N1 avian influenza given the outbreak in U.S. dairy cattle. *Lancet Reg Health Am* 2024; **35**: 100785.
3. Burrough ER, Magstadt DR, Petersen B *et al*. Highly pathogenic avian influenza A (H5N1) clade 2.3.4.4b virus infection in domestic dairy cattle and cats, United States, 2024. *Emerg Infect Dis* 2024; **30**: 1335–1343.

4. Shabil M, Khatib MN, Gaidhane S *et al*. Emerging threats in public health: H5N1 transmission from dairy cattle to humans. *New Microbes New Infect* 2024; **60–61**: 101429.
5. Branda F, Ciccozzi A, Romano C *et al*. Insights into avian influenza A (H5N1) events: epidemiological patterns and genetic analysis. *Infect Dis (Lond)* 2024; **56**: 678–681.
6. Uyeki TM, Milton S, Abdul Hamid C *et al*. Highly pathogenic avian influenza A (H5N1) virus infection in a dairy farm worker. *N Engl J Med* 2024; **390**: 2028–2029.
7. Ahrens AK, Jonsson SR, Svansson V *et al*. Iceland: an underestimated hub for the spread of high-pathogenicity avian influenza viruses in the North Atlantic. *J Gen Virol* 2024; **105**: e001985.
8. Rivetti AV Jr, Reischak D, de Oliveira CHS *et al*. Phylodynamics of avian influenza A (H5N1) viruses from outbreaks in Brazil. *Virus Res* 2024; **347**: 199415.
9. Paternina D, Herazo R, Oviedo M, Mattar S. Dramatic re-emergence of avian influenza in Colombia and Latin America. *Travel Med Infect Dis* 2024; **59**: 102711.
10. Kong D, He Y, Wang J *et al*. A single immunization with H5N1 virus-like particle vaccine protects chickens against divergent H5N1 influenza viruses and vaccine efficacy is determined by adjuvant and dosage. *Emerg Microbes Infect* 2024; **13**: 2287682.
11. Zhao Y, Chen P, Hu Y *et al*. Recombinant duck enteritis virus bearing the hemagglutinin genes of H5 and H7 influenza viruses is an ideal multivalent live vaccine in ducks. *Emerg Microbes Infect* 2024; **13**: 2284301.
12. Dashti F, Raisi A, Pourali G *et al*. A computational approach to design a multiepitope vaccine against H5N1 virus. *Virol J* 2024; **21**: 67.
13. Leong SL, Gras S, Grant EJ. Fighting flu: novel CD8$^+$ T-cell targets are required for future influenza vaccines. *Clin Transl Immunology* 2024; **13**: e1491.
14. Yap KL, Ada GL, McKenzie IF. Transfer of specific cytotoxic T lymphocytes protects mice inoculated with influenza virus. *Nature* 1978; **273**: 238–239.
15. Hamada H, Bassity E, Flies A *et al*. Multiple redundant effector mechanisms of CD8$^+$ T cells protect against influenza infection. *J Immunol* 2013; **190**: 296–306.
16. Taylor PM, Askonas BA. Influenza nucleoprotein-specific cytotoxic T-cell clones are protective *in vivo*. *Immunology* 1986; **58**: 417–420.
17. McMichael AJ, Gotch FM, Noble GR, Beare PA. Cytotoxic T-cell immunity to influenza. *N Engl J Med* 1983; **309**: 13–17.
18. Sridhar S, Begom S, Bermingham A *et al*. Cellular immune correlates of protection against symptomatic pandemic influenza. *Nat Med* 2013; **19**: 1305–1312.
19. Wang Z, Wan Y, Qiu C *et al*. Recovery from severe H7N9 disease is associated with diverse response mechanisms dominated by CD8$^+$ T cells. *Nat Commun* 2015; **6**: 6833.
20. Migueles SA, Sabbaghian MS, Shupert WL *et al*. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proc Natl Acad Sci USA* 2000; **97**: 2709–2714.
21. Lineburg KE, Grant EJ, Swaminathan S *et al*. CD8$^+$ T cells specific for an immunodominant SARS-CoV-2 nucleocapsid epitope cross-react with selective seasonal coronaviruses. *Immunity* 2021; **54**: 1055–1065.e5.
22. Augusto DG, Murdolo LD, Chatzileontiadou DSM *et al*. A common allele of HLA is associated with asymptomatic SARS-CoV-2 infection. *Nature* 2023; **620**: 128–136.
23. Lee LY, Ha DLA, Simmons C *et al*. Memory T cells established by seasonal human influenza A infection cross-react with avian influenza A (H5N1) in healthy individuals. *J Clin Invest* 2008; **118**: 3478–3490.
24. Quinones-Parra S, Grant E, Loh L *et al*. Preexisting CD8$^+$ T-cell immunity to the H7N9 influenza A virus varies across ethnicities. *Proc Natl Acad Sci USA* 2014; **111**: 1049–1054.
25. Kreijtz JH, de Mutsert G, van Baalen CA, Fouchier RA, Osterhaus AD, Rimmelzwaan GF. Cross-recognition of avian H5N1 influenza virus by human cytotoxic T-lymphocyte populations directed to human influenza A virus. *J Virol* 2008; **82**: 5161–5166.
26. Gras S, Kedzierski L, Valkenburg SA *et al*. Cross-reactive CD8$^+$ T-cell immunity between the pandemic H1N1-2009 and H1N1-1918 influenza A viruses. *Proc Natl Acad Sci USA* 2010; **107**: 12599–12604.
27. Sidney J, Kim A-R, de Vries RD *et al*. Targets of influenza Human T cell response are mostly conserved in H5N1. *bioRxiv*. 2024. https://doi.org/10.1101/2024.09.09.612060
28. GitHub – andersen-lab/avian-influenza: consensus sequences for U.S. H5N1 clade 2.3.4.4b. Available from: https://github.com/andersen-lab/avian-influenza.
29. Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics* 2007; **8**: 361.
30. NetMHCpan – 4.1. Available from: https://services.healthtech.dtu.dk/services/NetMHCpan-4.1/
31. Nguyen AT, Szeto C, Gras S. The pockets guide to HLA class I molecules. *Biochem Soc Trans* 2021; **49**: 2319–2331.
32. Szeto C, Lobos CA, Nguyen AT, Gras S. TCR recognition of peptide-MHC-I: rule makers and breakers. *Int J Mol Sci* 2020; **22**: 68.
33. Grant EJ, Josephs TM, Loh L *et al*. Broad CD8$^+$ T cell cross-recognition of distinct influenza A strains in humans. *Nat Commun* 2018; **9**: 5427.
34. Wu C, Zanker D, Valkenburg S *et al*. Systematic identification of immunodominant CD8$^+$ T-cell responses to influenza A virus in HLA-A2 individuals. *Proc Natl Acad Sci USA* 2011; **108**: 9178–9183.
35. Grant E, Wu C, Chan KF *et al*. Nucleoprotein of influenza A virus is a major target of immunodominant CD8$^+$ T-cell responses. *Immunol Cell Biol* 2013; **91**: 184–194.
36. Nguyen AT, Lau HMP, Sloane H *et al*. Homologous peptides derived from influenza A, B and C viruses induce variable CD8$^+$ T cell responses with cross-reactive potential. *Clin Transl Immunology* 2022; **11**: e1422.
37. Sant S, Quinones-Parra SM, Koutsakos M *et al*. HLA-B*27:05 alters immunodominance hierarchy of universal influenza-specific CD8$^+$ T cells. *PLoS Pathog* 2020; **16**: e1008714.
38. Poh CM, Zheng J, Channappanavar R *et al*. Multiplex screening assay for identifying cytotoxic CD8$^+$ T cell epitopes. *Front Immunol* 2020; **11**: 400.

39. van de Sandt CE, Kreijtz JH, de Mutsert G *et al*. Human cytotoxic T lymphocytes directed to seasonal influenza A viruses cross-react with the newly emerging H7N9 virus. *J Virol* 2014; **88**: 1684–1693.
40. Valkenburg SA, Josephs TM, Clemens EB *et al*. Molecular basis for universal HLA-A*0201-restricted $CD8^+$ T-cell immunity against influenza viruses. *Proc Natl Acad Sci USA* 2016; **113**: 4440–4445.
41. Grant EJ, Josephs TM, Valkenburg SA *et al*. Lack of heterologous cross-reactivity toward HLA-A*02:01 restricted viral epitopes is underpinned by distinct alphabetaT cell receptor signatures. *J Biol Chem* 2016; **291**: 24335–24351.
42. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 2003; **31**: 3784–3788.