Published in final edited form as:

J Immunol Methods. 2011 November 30; 374(1-2): 13–17. doi:10.1016/j.jim.2011.08.004.

Identification of potential human respiratory syncytial virus and metapneumovirus T cell epitopes using computational prediction and MHC binding assays

Michael T. Rock^{a,b}, Brett A. McKinney^e, Sandra M. Yoder^b, Catherine E. Prudom^{a,c}, David W. Wright^c, and James E. Crowe Jr.^{a,b,d,*}

^aDepartment of Pediatrics, Vanderbilt University Medical Center, Nashville, TN

^bVanderbilt Vaccine Center, Vanderbilt University Medical Center, Nashville, TN

^cDepartment of Chemistry, Vanderbilt University Medical Center, Nashville, TN

^dDepartment of Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN

^eDepartment of Mathematical and Computer Sciences, University of Tulsa and Laureate Institute for Brain Research, Tulsa, OK

Abstract

Human respiratory syncytial virus (RSV) and human metapneumovirus (MPV) are two of the most common causes of serious viral lower respiratory tract illness in humans. CD8⁺ T cells have been shown to be important in animal models and human clinical studies for the clearance of viral infection, and they may contribute in part to protection against severe disease during reinfections. Precise enumeration and accurate phenotyping of RSV- or MPV-specific CD8⁺ T cells in humans is currently limited by the relatively small number of T cell epitopes that have been mapped with accompanying identification of MHC restriction patterns. We sought to expand the number of potential RSV and MPV epitopes for use in clinical and translational studies by identifying an expanded set of MHC-binding peptides based on RSV and MPV wild-type virus strain protein sequences. We interrogated the full protein sequences of all 9 or 11 proteins of MPV or RSV respectively using four established epitope prediction algorithms for human HLA A*0101, A*0201, or B*0702 binding and attempted to synthesize the top-scoring 150-152 peptides for each of the two viruses. Synthesis resulted in 442 synthesized and soluble peptides of the 452 predicted epitopes for MPV or RSV. We then determined the binding of the synthetic peptides to recombinant human HLA A*0101, A*0201 or B*0702 molecules with the predicted restriction using a commercially available plate-based assay, iTopia. A total of 230 of the 442 peptides tested exhibited binding to the appropriate MHC molecule. The binding results suggested that existing algorithms for prediction of MHC A*0201 binding are particularly robust. The binding results also provided a large benchmarking data collection for comparison of new prediction algorithms.

Keywords

T-Lymphocytes; Immunologic techniques; Epitopes; T-Lymphocyte; Computational Biology; MHC binding peptide

*Corresponding author: James E. Crowe, Jr., MD, Vanderbilt University Medical Center, Vanderbilt Vaccine Center, T-2220 Medical Center North, 1161 21st Avenue South, Nashville, TN 37232-2905, USA, Tel.: (615) 343-8064, Fax: (615) 343-4456, james.crowe@vanderbilt.edu.

1. Introduction

RSV, a member of the Pneumovirus genus of the *Paramyxoviridae* family, causes relatively mild common cold symptoms in immunocompetent adults. However, RSV is the single most important cause of both severe lower respiratory tract infection during infancy and early childhood worldwide, and of hospitalization of infants in developed countries (Collins and Crowe, 2006). Nearly all children have had at least one RSV infection by 2 years of age and nearly one third of infants who have primary RSV infections develop lower respiratory tract infections. Furthermore, RSV lower respiratory tract illness in early childhood is an independent risk factor for the subsequent development of wheezing in children up to age 11 years (Stein et al., 1999). Human metapneumovirus (MPV) likewise is a major cause of lower respiratory tract illness in infants and children; it is often found as the second most common cause of lower respiratory tract illness (Williams et al., 2004). Mechanisms of immunity against disease caused by RSV and MPV are not fully understood, however most experts agree that CD8⁺ T cells are critical for resolution of established infection, and they may contribute to prevention of severe disease during reinfection (Collins and Crowe, 2006). A barrier to precise delineation of the number and phenotype of T cells in humans responding to RSV or MPV is the relatively small panel of T cell epitopes identified to date. Investigators have published a number of epitopes, typically identified using synthesis of overlapping peptides and testing of immune donor peripheral blood cells by interferon gamma ELISPOT testing (Goulder et al., 2000; Rock and Crowe, 2003; Venter et al., 2003).

In this study, we sought to expand the number of potential human MHC Class I restricted epitopes using a combination of computational prediction algorithms and *in vitro* MHC molecule binding assays. The large dataset of MHC binding assay results also served as the benchmarking dataset for the “Machine Learning in Immunology (MLI) competition” to compare the efficiency of new epitope prediction tools, which is described widely in this issue of the *Journal*.

2. Materials and methods

2.1. RSV and MPV sequences

The full-length viral genome sequences for RSV wild-type strain A2 (based on the parental virus sequence of *cp*-RSV, GenBank accession number AF035006) or MPV wild-type strain NL/00/17 type A2 (GenBank accession number FJ168779) were obtained from GenBank. The reading frames for these viral proteins were determined by others previously; we translated the GenBank nucleotide sequences into amino acid sequences of each of the previously described proteins encoded by viral genes using MacVector software version 11.0.

2.2. Algorithm predictions

To identify candidate peptide epitopes from RSV or MPV for synthesis, we entered each of the 11 RSV and 9 MPV protein sequences into four separate computer algorithms; (1) Artificial Neural Network (ANN; 9-mer peptides only) (Nielsen et al., 2003), (2) Average Relative Binding (ARB) (Bui et al., 2005), (3) Stabilized Matrix Method (SMM) (Peters and Sette, 2005), and (4) the BioInformatics and Molecular Analysis Section (BIMAS) (Parker et al., 1994). The ANN, ARB, and SMM algorithms were used with or without the influence of immuno-proteosomal cleavage predictions (Tenzer et al., 2005). In total, this analysis provided 7 scores for all 8,539 9-mer peptides and 5 scores for all 8,519 10-mer peptides from RSV and MPV.

We analyzed the potential for interaction with human MHC A*0101, A*0201, or B*0702 molecules. All algorithms were accessed via internet at the IEDB Analysis Resource

(www.immuneepitope.org; version 2006-01-01) and BIMAS (www.bimas.cit.nih.gov/molbio/hla_bind) websites. The scores for each of the algorithms for all peptides was aligned to compare the ANN, ARB, SMM (with and without proteosomal prediction influence), and BIMAS scores. We developed a searchable database for collection of the experimental data and the algorithm predictions using a commercial database software (Microsoft Access/Windows Operating System). We modified the database designations/headers in order to allow subsequent statistical analysis of the large datasets in SPSS. We used the database to construct and perform custom queries using relatively logical syntax that could retrieve comparisons, such as, what are the RSV peptides that score in the top peptides in ANN, ARB, SMM, and BIMAS algorithms. Each of the seven (9-mer peptides) or five (10-mer peptide) algorithm scores were weighted equally and we selected the top 150-152 average scores for each allele in the combined RSV/MPV dataset for synthesis and testing in MHC binding assays (Supplementary Tables 1-3).

2.3. Peptide synthesis

Peptides were synthesized on an Apex 396 Multiple Peptide Synthesizer (Advanced Chemtech) equipped with a 96-well reaction block capable of vortex mixing. Prefunctionalized wang resin was swollen in dichloromethane (Fisher) prior to synthesis. 9-Fluorenylmethoxycarbonyl (Fmoc) amino acids (Synpep) were coupled using O-benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluor-phosphate (HBTU; 5 eq. with respect to resin; Synpep), 1-hydroxybenzotriazole (HOBt; 5 eq.; SynPep), diisopropylethylamine (DIEA; 10 eq.; Advanced Chemtech) in N,N-dimethylformamide (DMF; Fisher). Fmoc protecting groups were removed with 20% piperidine in DMF. Peptides were cleaved with Reagent R (trifluoroacetic acid:thioanisole:ethanedithiol:anisole = 90:5:3:2 (v/v)) and desalted on a G-25 Sephadex column before QC with MALDI Mass Spectrometry (Applied Biosystems Voyager). Peptides that were not of suitable purity underwent a final purification on a C18 semi-preparative reverse phase HPLC column using water and acetonitrile gradients.

2.4. MHC binding assays

Purified peptides were tested for MHC binding using a commercial plate-based binding assay, iTopia (Coulter)(Wulf et al., 2009), per the manufacturer's instructions. The concept underlying the iTopia system and the procedures for conducting the assay are reviewed in detail in an accompanying article by Julian Montero in this issue of the *Journal*. Negative control peptides provided by the manufacturer were included in all MHC binding assays (the negative controls bound at 0% of the positive control).

3. Results

3.1. Algorithm predictions

Our goal was to identify the top approximately 150 predicted MHC class I (for HLA A*0101, A*0201, or B*0702) epitopes from either RSV or MPV. We began by running each of the 11 RSV and 9 MPV proteins through four prediction algorithms. The first algorithm we used was the BioInformatics and Molecular Analysis Section (BIMAS), available at www.bimas.cit.nih.gov/molbio/hla_bind, which provided the estimated half-time disassociation of each possible 9- and 10-mer peptide. Next we utilized three T-cell epitope prediction algorithms available through the IEDB Analysis Resource website (version 2006-01-01). Each algorithm predicted IC₅₀ [nM] scores for all possible 9-mer peptides (ANN, ARB, and SMM) or 10-mer peptides (ARB and SMM). Finally, we combined the ANN, ARB, and SMM algorithms with immuno-proteosomal cleavage prediction methods to provide an overall score calculated by combining the proteosomal cleavage, TAP transport and MHC binding predictions. In total, this provided 7 scores for

each possible 9-mer (8,539 peptides) and 5 scores for each possible 10-mer (8,519 peptides) from RSV and MPV. In general, the scores for individual peptides corresponded reasonably well between the four algorithms. We combined each of the algorithm scores for all possible 9-mer and 10-mer peptides and selected the top 150-152 scoring peptides for each MHC molecule, regardless of the virus or viral protein source. A complete listing of the algorithm predictions is shown in Supplementary Tables 1-3, available online.

3.2 Peptide synthesis

We attempted to synthesize the top scoring 150-152 individual peptides for each of the three MHC class I alleles, with processing to high purity. We were successful in synthesizing 442 of the 452 peptides selected. A number of the predicted peptides posed significant challenges to efficient synthesis or formulation into a solution for testing. Most commonly, the problem was one of extreme hydrophobicity. We synthesized 149 peptides for A*0101 binding, 145 peptides for A*0201 binding, and 148 peptides for B*0702 binding.

3.3 MHC binding results

The binding results of synthetic peptides were determined for binding to the appropriate MHC molecule, HLA A*0101, A*0201, or B*0702. We used control peptides, EVDPIGHLY for HLA-A*0101 (from MAGE 3 gene (melanoma/tumor antigen) (Gaugler et al., 1994), FLPSDFFPSV for HLA-A*0201 (from hepatitis B virus nucleocapsid protein) (Bertoletti et al., 1993) or GPRLGVRAT for HLA-B*0702 (from hepatitis C virus) (Koziel et al., 1995), as provided by the manufacturer in the iTopia system to determine a positive signal for binding in the assay. The positive control peptides had been selected for optimal affinity and stability. Then, we compared the signal for binding of the RSV or MPV peptides to the same MHC molecule. The assays were run in duplicate, and the results were combined and the results reported as % of the positive control. The correspondence of values between the duplicate wells was quite high, revealing that the intra-run reproducibility was high. Intra-assay percent coefficients of variation for the peptide binding assays were on average 2.7 for HLA A*0101, 2.7 for HLA A*0201, and 1.8 for HLA B*0702.

The data revealed interesting patterns (Table 1 and 2). The complete binding results are shown in Supplementary Tables 1-3, available online. The binding signals exhibited a bimodal distribution, with high binding or low binding signals. Based on the distribution, we chose to use a threshold of 30% binding compared to the positive control for that allele to select those peptides considered as binders. For HLA A*0101, 39 of 149 (26%) predicted epitopes exhibited binding (46% in the top 25 predicted were binding). For HLA A*0201, 138 of 145 (95%) predicted epitopes exhibited binding. For HLA B*0702, 53 of 148 (36%) predicted epitopes exhibited binding (28% in the top 25 predicted were binding). In total, 230 of the 442 peptides tested (52%) exhibited MHC binding.

4. Discussion

The identification of human T cell epitopes in major human viral pathogens is of interest for facilitating studies to enumerate circulating T cells following viral infection of vaccination. Epitope specific studies using peptides or MHC-peptide tetramers allow investigators to perform precise phenotyping assays of virus-specific cells using polychromatic flow cytometry. It is possible to synthesize all of the potential peptides in a virus for testing for reactivity in assays with human immune donor cells, but the logistics of this approach for RSV and MPV epitope discovery are not favorable for several reasons. First, the peptide cost is high, well over a half of a million U.S. dollars in the case of these viruses. Second, the amount of blood needed for performing the assays is excessive, potentially requiring

leukoapheresis of donors for testing complete peptide sets. Third, the frequency of peptide-specific T cells in the circulation is quite low in otherwise healthy donors, posing a challenge to large-scale screening. Circulating epitope-specific T cell numbers may be elevated in subjects with acute or recent RSV or MPV infection, however most subjects with easily identified RSV or MPV infection in the clinical setting are infants. Instead, alternative methods have been employed to narrow the number of possible epitopes prior to testing with human blood samples. Most commonly, predictive computational algorithms are used to identify peptides with potentially favorable processing and binding characteristics.

Here, we used several established algorithms to identify a candidate set of 452 RSV or MPV peptides that were predicted to bind to one of three common human HLA molecules, then tested the binding of the synthetic peptides in a commercially available MHC binding assay. We used a consensus strategy of scoring using multiple algorithms to identify the potential binding peptides. In this type of study, it is not uncommon to encounter predicted peptides that are difficult to produce synthetically, usually because of extreme hydrophobicity. Indeed, of the 452 top-scoring predicted MHC binding peptides, we were successful in synthesizing 442 (98%). It was not possible to test whether the remaining can be recognized as epitopes using these *in vitro* methods.

Interestingly, the accuracy of the algorithms for prediction of MHC binding varied among the alleles. The HLA A*0201 algorithms were highly effective, with 95% of peptides exhibiting binding to the predicted allele. In contrast, only about a quarter to a third of the predicted epitopes for HLA A*0101 or B*0702 bound to the predicted MHC molecule. Probably this reflects the greater maturity of the work on HLA A*0201, which was targeted earliest for development because HLA A*0201 is the most common MHC type in typical human volunteer populations. Benchmarking datasets containing MHC binding data for development and testing of new prediction algorithms are needed. We provided the data on binding of these peptides to the “Machine Learning in Immunology (MLI) competition”, the results of which are widely discussed in this issue of the *Journal*.

The *in vitro* method we used for determining MHC binding is somewhat complex from a conceptual standpoint, however is straightforward in practice. The assay uses mixture of a biotinylated unfolded recombinant MHC molecule, β 2-microglobulin, and test peptide with incubation in a streptavidin-coated plate. If the test peptide stabilizes the MHC and β 2-microglobulin association, proper folding of the MHC molecule can be recognized with secondary addition of a conjugated antibody that recognizes the folded conformation in concert with a substrate. The readout of binding is conducted in a blinded fashion using a conventional plate reader. We found replicate runs of the assay to yield reproducible results.

All of the results in this paper suggest *potential* human T cell epitopes in RSV or MPV, as evidenced by MHC binding. These epitopes have not been validated in humans yet. Reasons that some of these peptides might have the potential to bind an appropriate MHC molecule but not serve as human T cell epitopes include lack of efficient processing of the peptide from longer polypeptides, lack of loading onto MHC molecules, lack of an appropriate T cell receptor to recognize the MHC-peptide complex, anergy of T cells recognizing the MHC-peptide complex, and others. Nevertheless, the data provide a large but focused list of peptides that could be tested in humans much more efficiently than a panel of all peptides of 9 to 10 amino length in the viral proteins.

5. Summary

In conclusion, we generated MHC binding data for 442 peptides derived from RSV and MPV protein sequences, yielding a robust dataset for benchmarking new T cell epitope

algorithms. The data revealed that existing Class I MHC prediction algorithms work reasonably well for predicting MHC binding, with the HLA A*0201 algorithms predictive in about 90% of cases. We identified 220 peptides from RSV or MPV that bind to HLA A*0101, A*0201, or B*0702 that can be used in future human studies of T cell response to these major human pathogens.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by R01 AI-53222, NIAID, NIH. We thank Joshua D. Swartz and Ryan D. Rutledge for excellent technical support. JEC is a Burroughs Wellcome Fund Clinical Scientist in Translational Research.

References

- Bertoletti A, Chisari FV, Penna A, Guilhot S, Galati L, Missale G, Fowler P, Schlicht HJ, Vitiello A, Chesnut RC, et al. Definition of a minimal optimal cytotoxic T-cell epitope within the hepatitis B virus nucleocapsid protein. *J Virol*. 1993; 67:2376–80. [PubMed: 7680391]
- Bui HH, Sidney J, Peters B, Sathiamurthy M, Sinichi A, Purton KA, Mothe BR, Chisari FV, Watkins DI, Sette A. Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. *Immunogenetics*. 2005; 57:304–14. [PubMed: 15868141]
- Collins, PL.; Crowe, JE, Jr. Respiratory syncytial virus and metapneumoviruses. In: Knipe, DM.; Howley, PM., editors. *Fields Virology*. Vol. Chapter 46. Lippincott, Williams & Wilkins; Philadelphia: 2006. p. 1601-1646.
- Gaugler B, Van den Eynde B, van der Bruggen P, Romero P, Gaforio JJ, De Plaen E, Lethe B, Brasseur F, Boon T. Human gene MAGE-3 codes for an antigen recognized on a melanoma by autologous cytolytic T lymphocytes. *J Exp Med*. 1994; 179:921–30. [PubMed: 8113684]
- Goulder PJ, Lechner F, Klenerman P, McIntosh K, Walker BD. Characterization of a novel respiratory syncytial virus-specific human cytotoxic T-lymphocyte epitope. *J Virol*. 2000; 74:7694–7. [PubMed: 10906229]
- Koziel MJ, Dudley D, Afdhal N, Grakoui A, Rice CM, Choo QL, Houghton M, Walker BD. HLA class I-restricted cytotoxic T lymphocytes specific for hepatitis C virus. Identification of multiple epitopes and characterization of patterns of cytokine release. *J Clin Invest*. 1995; 96:2311–21. [PubMed: 7593618]
- Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, Brunak S, Lund O. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci*. 2003; 12:1007–17. [PubMed: 12717023]
- Parker KC, Bednarek MA, Coligan JE. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J Immunol*. 1994; 152:163–75. [PubMed: 8254189]
- Peters B, Sette A. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics*. 2005; 6:132. [PubMed: 15927070]
- Rock MT, Crowe JE Jr. Identification of a novel human leucocyte antigen-A*01-restricted cytotoxic T-lymphocyte epitope in the respiratory syncytial virus fusion protein. *Immunology*. 2003; 108:474–80. [PubMed: 12667209]
- Stein RT, Sherrill D, Morgan WJ, Holberg CJ, Halonen M, Taussig LM, Wright AL, Martinez FD. Respiratory syncytial virus in early life and risk of wheeze and allergy by age 13 years. *Lancet*. 1999; 354:541–5. [PubMed: 10470697]
- Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, Kloetzel PM, Rammensee HG, Schild H, Holzthutter HG. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci*. 2005; 62:1025–37. [PubMed: 15868101]

- Venter M, Rock M, Puren AJ, Tiemessen CT, Crowe JE Jr. Respiratory syncytial virus nucleoprotein-specific cytotoxic T-cell epitopes in a South African population of diverse HLA types are conserved in circulating field strains. *J Virol.* 2003; 77:7319–29. [PubMed: 12805430]
- Williams JV, Harris PA, Tollefson SJ, Halburnt-Rush LL, Pingsterhaus JM, Edwards KM, Wright PF, Crowe JE Jr. Human metapneumovirus and lower respiratory tract disease in otherwise healthy infants and children. *N Engl J Med.* 2004; 350:443–50. [PubMed: 14749452]
- Wulf M, Hoehn P, Trinder P. Identification of human MHC class I binding peptides using the iTOPIA-epitope discovery system. *Methods Mol Biol.* 2009; 524:361–7. [PubMed: 19377958]

Abbreviations used in this paper

RSV	respiratory syncytial virus
MPV	metapneumovirus
MHC	major histocompatibility complex
HLA	human leukocyte antigen
PEG	polyethelene glycol

Table 1

Percent of predicted peptides tested that bound to the appropriate MHC molecule

HLA molecule	Number of predicted peptides tested	% tested that bound to the predicted MHC molecule (number binding)	% of the <i>top 25</i> ranked peptides that bound
A*0101	149	26 (39)	46
A*0201	145	95 (138)	72
B*0702	148	36 (53)	28

Table 2

Number of 9-mer and 10-mer peptides for each protein that bound to the appropriate MHC molecule (number of peptides tested)

Protein	A*0101		A*0201		B*0702	
	9-mer	10-mer	9-mer	10-mer	9-mer	10-mer
MPV-F	1 (3)	1 (3)	2 (2)	5 (5)	2 (3)	1 (3)
MPV-G	1 (3)	1 (4)	NA	3 (3)	1 (2)	0 (2)
MPV-L	5 (19)	2 (8)	18 (20)	17 (19)	9 (19)	1 (15)
MPV-M	0 (4)	0 (1)	2 (2)	NA	3 (4)	2 (4)
MPV-M2-1	1 (3)	1 (2)	3 (3)	1 (1)	1 (4)	NA
MPV-M2-2	0 (1)	NA	NA	1 (1)	1 (1)	NA
MPV-N	2 (6)	NA	2 (2)	1 (1)	6 (11)	0 (3)
MPV-P	1 (1)	NA	1 (1)	NA	1 (1)	1 (3)
MPV-SH	NA	0 (2)	1 (1)	2 (2)	NA	0 (2)
RSV-F	3 (7)	2 (5)	6 (6)	6 (6)	2 (2)	0 (2)
RSV-G	0 (1)	NA	3 (4)	NA	0 (2)	0 (4)
RSV-L	7 (33)	5 (19)	23 (24)	19 (19)	9 (20)	3 (9)
RSV-M	1 (3)	1 (4)	NA	NA	0 (1)	1 (8)
RSV-M2	NA	0 (1)	1 (1)	2 (2)	NA	0 (2)
RSV-M2orf	0 (4)	1 (1)	1 (1)	NA	1 (2)	NA
RSV-N	1 (5)	NA	3 (4)	1 (1)	3 (7)	0 (1)
RSV-NS1	1 (1)	0 (1)	3 (3)	1 (1)	NA	NA
RSV-NS2	1 (2)	NA	3 (3)	2 (2)	2 (4)	1 (2)
RSV-P	0 (1)	NA	1 (1)	1 (1)	1 (3)	0 (1)
RSV-SH	NA	0 (1)	3 (3)	NA	1 (1)	NA