

2011

# A Comparison of Two Methods for T Cell Epitope Mapping: “Cell Free” In Vitro Versus Immunoinformatics

Timothy J. Messitt

Frances Terry

*See next page for additional authors*

Creative Commons License



This work is licensed under a [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/).

Follow this and additional works at: [https://digitalcommons.uri.edu/immunology\\_facpubs](https://digitalcommons.uri.edu/immunology_facpubs)

Terms of Use

All rights reserved under copyright.

---

## Citation/Publisher Attribution

Messitt, T. J., Terry, F., Moise, L., Martin, W., & De Groot, A. S. (2011). A comparison of two methods for T cell epitope mapping: “cell free” in vitro versus immunoinformatics. *Immunome Research*, 7(2): e6.

Available at: <http://dx.doi.org/10.4172/1745-7580.1000045>

This Article is brought to you for free and open access by the Institute for Immunology and Informatics (iCubed) at DigitalCommons@URI. It has been accepted for inclusion in Institute for Immunology and Informatics Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact [digitalcommons@etal.uri.edu](mailto:digitalcommons@etal.uri.edu).

---

**Authors**

Timothy J. Messitt, Frances Terry, Leonard Moise, William Martin, and Anne S. De Groot



RESEARCH

Open Access

# A comparison of two methods for T cell epitope mapping: “cell free” in vitro versus immunoinformatics

Timothy J. Messitt<sup>1</sup>, Frances Terry<sup>1</sup>, Leonard Moise<sup>1,2</sup>, William Martin<sup>1</sup>, Anne S. De Groot<sup>1,2,3 §</sup>

## Abstract

### Background

Methods for identifying physiologically relevant T-cell epitopes are critically important for development of vaccines and the design of therapeutic proteins. As the number of proteins that are being evaluated for putative immunogenicity expands, rapid and accurate tools are in great demand. Several methods to identify T-cell epitopes have been developed, the most recent of which is a cell free system consisting of a minimal set of proteases incubated with HLA DRB1\*0101, HLA-DM and whole antigen. Isolation and sequencing of the HLA bound peptides using mass spectrometry allows for the prospective identification of immunodominant T-cell epitopes.

### Results

We present here, a comparison of this cell free in vitro antigen processing system to an immunoinformatics approach using the EpiMatrix algorithm. Our comparison reveals that in addition to identifying a similar set of epitopes to the cell-free system, the immunoinformatics approach prospectively identifies more HLA-DRB1\*0101 epitopes and can simultaneously analyze multiple HLA alleles.

### Conclusions

Although the cell-free system incorporates antigen processing and MHC binding, the immunoinformatics approach identifies many validated epitopes with a very high degree of accuracy and can be performed much faster with far fewer resources.

## Background

Methods for the prospective identification of physiologically relevant T-cell epitopes are critically important for development of vaccines and for the design of therapeutic proteins. A cell free system (CFS) for prospectively identifying T-cell epitopes from whole antigens was recently described and applied to the identification of influenza epitopes [1]. As described by Hartman et al. in their publication, CFS epitope mapping was performed by pre-incubating whole antigens with HLA-DRB1\*0101 and HLA-DM, and then exposing the mixture of antigen and HLA DR/DM to a minimal set of proteases, followed by isolation and sequencing of the HLA-bound peptides using mass spectrometry. The CFS was initially validated using two model antigens (HA1 from influenza A/Texas/1/77 and type II collagen) as positive controls and then applied prospectively for the discovery of new HLA-DRB1\*0101 immunodominant epitopes from a recombinant liver-stage antigen of malaria falciparum (LSA-NRC) and HA1 from H5N1 influenza (Viet Nam).

The publication of the CFS method provided an opportunity for comparing a purely immunoinformatics approach

based entirely on MHC binding affinity (EpiMatrix) to an in vitro system that involves both antigen processing and presentation [2]. We hypothesized that predicted MHC binding (as performed in silico) would provide results that were at least equivalent to the more laborious CFS approach. As the identification of T cell epitopes using the CFS approach requires a significant amount of laboratory effort, reagents, and specific expertise in the use of MALDI mass-spectrometry, the immunoinformatics approach might, in addition, offer significant time and cost savings. As is described here, our detailed comparison reveals that the immunoinformatics method correctly identified four of the six epitopes identified by the CFS method, at lower cost and with greater time efficiency, and, in addition, identified other potential epitopes that appear to have been missed by the CFS. Neither of the two CFS epitopes that were missed by EpiMatrix were validated in follow up assays. In the brief report below, we provide a detailed comparison of the *in silico* approach using EpiMatrix and the CFS approach, in terms of epitopes identified and the relative speed, effort required and cost of the two methods.

## Results

### CFS reductionist method

The CFS approach to finding immunodominant epitopes, as published in reference 1, is described here for comparison with the EpiMatrix method. The cell free system (CFS) is restricted to evaluations of a single HLA at a

1 EpiVax, Inc., Providence, RI, USA

2 Institute for Immunology and Informatics, University of Rhode Island, Providence, RI, USA

3 The Warren Alpert Medical School of Brown University, Providence, RI, USA

§Corresponding author

Email addresses:

TJM: tmessitt@epivax.com

FT: fterry@epivax.com

LM: lmoise@epivax.com

WM: martinb@epivax.com

ASD: annied@epivax.com





Code (Fig.1)	AA Sequence	Core EpiMatrix Epitope	DRB1*0101 Z-Score	IEDB Reference ID
C-VN45	SNGNFIAPEYAYKIVKKGDS	IAPEYAYKI	1.53	1009685**
C-LSA11	YDNFQDEENIGIYK	FQDEENIGI	1.58	1002410**
I/C-CII5	QTGEPGIAGFKGEGQKPGKEGPGAGVQGGAPGPAG	FKGEGQKPGK	2.66	1007108*
I/C-PR1	AINSSLPYQNIHPVTIGCEPK	YQNIHPVTI	2.72	1000157†
I/C- <i>Tex1</i>	SRGACPKYVQKNTLKLATGMRK	YVKQNTLKL	3.06	1000083*
I/C-LSA2	EDITFMKLGSGSPHHHH	FMKLGSGGS	2.62	1002410**
I-CII1	YRSQKTSRL	YRSQKTSRL	3.14	N/A
I-CII2	FLRLLSTEG	FLRLLSTEG	2.98	N/A
I-CII3	FTGLQLPLG	FTGLQLPLG	2.75	1014795**
I-PR2	YQENAYVS	YQENAYVS	2.54	1000157†
I-PR3	WTLLKPGDT	WTLLKPGDT	2.53	1000157†
I-VN1	FHHIHLTI	FHHIHLTI	2.76	1018856**
I-VN2	LKHLSRIN	LKHLSRIN	2.74	1018856**
I-VN3	YIVEKANPV	YIVEKANPV	2.64	1009685*
I-LSA1	FKSLRLNLG	FKSLRLNLG	3.09	1002322**
I-LSA3	IKSNLRS GS	IKSNLRS GS	2.03	N/A

**Table 1 - Epitope Comparison**

Summary of epitopes identified by EpiMatrix and the CFS. In the case of I/C-PR1, I/C-*Tex1*, and C-VN45, multiple overlapping peptides were eluted in the CFS; those peptides sharing the same core 9-mer identified by EpiMatrix are considered one epitope. Variable flanking residues are indicated in grey text. Refer to Figure 1 for nomenclature and annotation.

As compared to the CFS, the EpiMatrix immunoinformatics approach identified many more prospective epitopes for each of the antigens. Performing the analysis as described here, a total of 13 epitopes were identified based on their HLA DRB1\*0101 score: these included the three highest-scoring epitopes for each of four antigens that were evaluated by the CFS and the one high-scoring HLA DRB1\*0101 epitope from A/Texas (only one published epitope was included in the rHA protein tested in the CFS). Of these 13 epitopes, eight (62%) were previously confirmed as human T cell epitopes, according to IEDB; validation is not yet recorded the remaining five epitopes.

#### Detailed comparison of CFS and EpiMatrix results

Epitope sequences identified by both the CFS and EpiMatrix method were cross-referenced against the Immune Epitope Database (IEDB) of published T cell epitopes and MHC ligands using the substring parameter. Further, all T cell responses and MHC binding results available on IEDB were compiled for each input antigen. If IEDB is taken as the reference standard for validation of epitopes predicted by either method, EpiMatrix prospectively identified five more ‘validated’ epitopes that the CFS method did not identify. CFS identified two epitopes that were not identified by EpiMatrix, and although these are published in IEDB, they have been published for alternative alleles. If more stringent criteria are applied, (restriction by HLA DRB1\*0101 and publication in IEDB), EpiMatrix correctly identified three of the three (100%) HLA DRB1\*0101-restricted epitopes. This significant reduction is due in part to the limited HLA-restriction references available on IEDB; of the four test antigens, only H5N1 Vietnam had references qualified as HLA-DRB1\*0101-restricted. As we will discuss in greater detail below, the single HLA DRB1\*0101-restricted epitope prospectively identified by the CFS (I/C-CII5), was only validated as a result of querying the EpiMatrix-identified core.

**Influenza PR antigen.** The single PR epitope that was identified by both the CFS and EpiMatrix, I/C-PR1, was ranked first of all PR-derived HLA DRB1\*0101 epitopes

by EpiMatrix, and was extracted from the same amino acid locus (306-318) as the influenza HA epitope control peptide (I/C-*Tex1*) that was fused to the A/PR/8/34 antigen. The two epitopes also identified by EpiMatrix have high DRB1\*0101 scores (I/C-PR1 and I/C-*Tex1*) and contain an epitope bar (or EpiBar) [6,7], a feature that is often present in promiscuous epitopes [5], and that has been associated with immunogenicity in human studies [8,9]. This important feature of promiscuous, immunogenic epitopes is not detectable using the CFS. EpiMatrix also identified two additional epitopes in influenza A/PR/8/34 HA (Table 1).

**Influenza A Viet Nam.** When the full length HA1 from H5N1 A/Vietnam/1203/2004 was analyzed in the CFS, two unique peptide species were identified; both shared the same core sequence (C-VN45). The authors selected the shorter of these two peptides (HA<sub>259-274</sub>) to test for immunodominance in T cell proliferation assays, cytokine production assays, and tetramer staining assays. EpiMatrix analysis shows that the shorter HA most likely has a truncated terminal HLA DRB1\*0101 HLA binding motif. The peptide only has one EpiBar at the C-terminal end of the sequence and the N-terminal end of the peptide, which is devoid of predicted HLA binding motifs, would interfere with binding, *in vitro*, and with immunogenicity. The properly centered peptide would have been identified prospectively using EpiMatrix.

Three other epitopes were predicted by EpiMatrix in the same protein to be better HLA DRB1\*0101 binders, but they were not identified by the CFS. All three have been experimentally validated and published; only one was confirmed for HLA DRB1\*0101. Based on the fact that the third-ranked EpiMatrix epitope is published and found to be an HLA DRB1\*0101 epitope in IEDB, the other two (ranked 1 and 2) are equally likely if not more likely to be HLA DRB1\*0101-restricted, however this would have to be tested prospectively in T cell assays as was done for the CFS epitopes.

**Collagen.** It is notable that the single collagen epitope (I/C-CII5) identified by the CFS was also confirmed by EpiMatrix. It was the 5th ranked peptide for HLA DRB1\*0101, which is not unexpected considering the greater length of the CII sequence compared to other proteins examined in the CFS. The peptide has a

maximum DR1 EpiMatrix score of 2.66, putting it in the top 1% of peptides expected to bind to DR1, a rank that would have normally led to selection by EpiMatrix. The sequence of the CFS eluted peptide CII<sub>273-305</sub> contains the same epitope that is predicted by EpiMatrix, however its sequence is elongated and the probable HLA binding motif is almost obscured by the extended flanks that were eluted with the core sequence. When queried against the IEDB, the CFS-eluted sequence returned no results, whereas the EpiMatrix-identified core sequence was published as a DRB1\*0101-restricted epitope. In other words, immunoinformatics analysis using EpiMatrix can rapidly and accurately identify a core HLA binding epitope, whereas the CFS does not. EpiMatrix identified three additional epitopes in CII that are likely to be immunogenic (ranked 1, 2, and 3). One was previously reported in IEDB but not confirmed in HLA transgenic mice nor in human T cell assays.

**Malaria LSA.** Finally, the single LSA epitope identified by both the Immunoinformatics approach and the CFS method contains a strong EpiMatrix-scoring peptide, and was the only LSA epitope to be validated in T cell assays. The weaker LSA epitope (C-LSA11) was also identified by the immunoinformatics approach, but the EpiMatrix score was slightly below the normal cut off for selection. Two additional epitopes, I-LSA1 and I-LSA3 were identified using EpiMatrix, of which one (I-LSA1) is found in IEDB for an unreported Class II HLA allele.

## Discussion

The CFS method described by Hartman et al. is a novel approach for identification of T cell epitopes that involves a significant amount of effort, reagents, and highly technical expertise in the use of MALDI mass-spectrometry. One advantage of CFS is that the method could, in theory, be applied to additional antigens and alleles for which immunoinformatics approaches do not yet exist. In comparison to the immunoinformatics approach, EpiMatrix, it appears to be both time- and cost-intensive. Many of the same epitopes are identified using the two approaches. Additionally, previously validated epitopes are identified using EpiMatrix that appear to be missed by the CFS (when comparing CFS and EpiMatrix results with the IEDB database of published and validated epitopes), as are several epitopes that appear likely to be immunogenic based on their rank and EpiMatrix score. Thus antigen-processing as performed using the CFS may have destroyed peptides that might otherwise have been presented by DR1. Two epitopes were identified using the CFS that would not have been selected by EpiMatrix. Neither of these epitopes would qualify as high quality targets based on information obtained from IEDB.

Epitopes identified using the CFS were not compared to epitopes that have already been published. Instead, *in vivo* methods were used to validate sequences. Epitopes were considered to be validated if, following immunization of HLA-DRB1\*0101 mice with whole antigen,

T cells were induced to proliferate, cytokine production was increased, or tetramer-stained cells could be detected, or some combination of the three. These methods are useful for validation regardless of the epitope selection method – and have been used for the validation of T cell epitopes identified using EpiMatrix (see references 6 and 9).

The CFS may provide some additional information on epitopes that are processed (two were identified in the CFS that were not identified using EpiMatrix) and may have application to organisms with little to no information on peptide-MHC interactions. A much more extensive (and perhaps cost-prohibitive) comparison would have to be performed, prospectively validating the epitopes discovered using EpiMatrix alongside epitopes discovered using the CFS, before either method is proven to be superior to the other. At the very least, they appear to be somewhat complementary.

It should be mentioned that there are a range of epitope prediction tools available, however each tool may provide slightly different results [3]. For example, when analyzing the sequences from H5N1 Vietnam and rLSA-1, both of which contain CFS-only epitopes, (Figure 1) with NetMHCII [10], an alternative epitope prediction tool, NetMHCII predicted the H5N1 Vietnam epitope but the rLSA-1 epitope was still not predicted. NetMHCII predicted similar epitopes to EpiMatrix for the proteins included in this analysis (data not shown) with the best scoring EpiMatrix epitopes centered within the sequences all of the strong binders predicted by NetMHCII. NetMHCII does not further delineate the optimal epitope when using its default peptide length of 15 amino acids (<http://www.cbs.dtu.dk/services/NetMHCII/>).

Immunoinformatics provides a rapid means of analyzing protein sequences for multiple HLA alleles at the same time. EpiMatrix prospectively identified more HLA-DRB1\*0101 epitopes than the CFS, of which almost all are present in IEDB. In addition to providing an analysis of multiple alleles at once (not just HLA DRB1\*0101) and highlighting promiscuous epitopes (data not shown, references 2 and 3 provide examples), the immunoinformatics approach can be performed at a fraction of the time and cost (one day, no reagents, versus multiple weeks, technician time, *in vitro* reagents, and MALDI mass spectrometer costs). Thus immunoinformatics approaches with algorithms such as EpiMatrix provide a clear cost and time advantage over CFS.

As the number of proteins that are being evaluated for putative immunogenicity expands, rapid, inexpensive and accurate tools are in great demand. The time and effort involved in prospectively identifying peptides from a single protein for a single allele were not reported for this study. Based on rough estimates for the costs, the immunoinformatics method can be performed in less than 1/20th of the time and for 1/100th of the cost of the cell free method. If one is to compare the cost of equipment that would be necessary for these two methods (a single internet accessible computer vs. a fully functional lab with mass spectrometry and protein production

capabilities), this gap grows exponentially wider.

## Conclusions

Adding antigen-processing to epitope selection in vitro does not improve prospective class II antigen identification. In this comparison of two methods for T cell epitope identification, the in vitro antigen-processing and MHC-elution method (CFS) obscured two of three externally validated, HLA-DRB1\*0101-restricted epitopes, whereas the immunoinformatics approach (EpiMatrix) correctly identified all published epitopes restricted by DRB1\*0101, and prospectively identified many additional epitopes that have yet to be validated. While the CFS approach may enable identification of epitopes from MHC for which there is little to no published information, the advantages of the immunoinformatics method, which include a very high degree of accuracy, high throughput, rapidity, and low cost, are clear.

## Methods

### Immunoinformatics

Proteins analyzed by EpiMatrix are parsed into 9-mers that overlap each other by 8 amino acids. Each of these 9-mers is then scored individually (and simultaneously) for predicted binding affinity against a panel of eight common Class II HLA alleles (DRB1\*0101, \*0301, \*0401, \*0701, \*0801, \*1101, \*1301, and \*1501) that collectively cover over 95% of the human population [11]. EpiMatrix scores range from approximately -3 to +3. Those 9-mer peptides that score 1.64 or above are considered to be potential epitopes. The predictions have been benchmarked against the 'gold standard' set of published epitopes, performing as well or better than other epitope mapping tools [3]. Typically, 5% of all nine-mer frames in a given protein score above 1.64; these are considered to have a significant chance of binding to HLA molecules with moderate to high affinity and to have a significant chance of being presented on the surface of APCs [3].

### Competing interests

EpiVax Inc. is a for-profit biotechnology company and all of the authors have a financial interest in or are employed by the company.

### Authors' contributions

TJM assisted in writing, editing and provided conceptual input toward this publication. FT and WM obtained the raw material for the research, determined the appropriate method for the comparison, performed the

immunoinformatics analysis and assisted in writing and editing the document. LM analyzed data, provided conceptual input and assisted in editing. ASD elaborated the hypothesis, directed the research, assisted with the writing and editing of the manuscript and provided conceptual input toward the publication.

## References

1. Hartman, IZ, Kim, A, Cotter, RJ, Walter, K, Dalai, SK, Boronina, T, Griffith, W, Lanar, DE, Schwenk, R, Krzyzch, U, Cole, RN, Sadegh-Nasseri, S. **A reductionist cell-free major histocompatibility complex class II antigen processing system identifies immunodominant epitopes.** *Nature Medicine*, 2010, **16**(11):1333-1341.
2. De Groot AS, Jesdale BM, Szu E, Schafer JR, Chicz RM, Deocampo G. **An interactive Web site providing major histocompatibility ligand predictions: application to HIV research.** *AIDS Res Hum Retroviruses.*, 1997, **13**, 529-31.
3. De Groot AS, Martin, WD. **Reducing risk, improving outcomes: Bioengineering less immunogenic protein therapeutics.** 2009, *Clin. Immunol.*, **131**:189-201.
4. Steffen C, Sanger L, Menzel, J. **Demonstration of antibodies to denatured type I and type II collagen in juvenile rheumatoid arthritis, Still's syndrome and controls by [14C] collagen radioimmunoassay.** *Scand J Rheumatol.*, 1980 **9** (2):69-76.
5. Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, Damle R, Sette A, Peters B. **The immune epitope database 2.0.** *Nucleic Acids Res.*, 2010, **38**: D854-62.
6. De Groot AS, Ardito M, McClaine E, Moise, L, Martin W. **Immunoinformatic comparison of T-cell epitopes contained in novel swine-origin influenza A (H1N1) virus with epitopes in 2008-2009 conventional influenza vaccine.** *Vaccine*, 2009, **27**: 5740-7.
7. Weber CA, Mehta, PJ, Ardito, M, Moise, L, Martin W, De Groot AS. **T cell epitope: friend or foe? Immunogenicity of biologics in context.** *Adv. Drug Deliv. Rev.*, 2009, **61**: 965-976.
8. Moss SF, Moise L, Lee, DS, Kim, W, Zhang S, Lee J, Rogers, AB, Martin, W, De Groot AS. **HelicoVax: Epitope-based therapeutic Helicobacter pylori vaccination in a mouse model.** *Vaccine*, 2011, **29**(11): 2085-91.
9. Moise L, Buller RM, Schriewer J, Lee J, Frey SE, Weiner DB, Martin, W, De Groot AS. **VennVax, a DNA-prime, peptide-boost multi-T-cell epitope poxvirus vaccine induces protective immunity against vaccinia infection by T cell response alone.** *Vaccine*, 2011 **29** (3):501-11.
10. Nielsen, M, Justesen, S., Lund, O, Lundegaard, C, Buus, S. **NetMHCIIpan-2.0- Improved pan-specific HLA-DR predictions using a novel concurrent alignment and weight optimization training procedure.** *Immunome Research*, 2010 Nov **13**:6:9.
11. Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, Kubo RT, Chestnut RW, Grey HM, Sette A: **Several common HLA-DR types share largely overlapping peptide binding repertoires.** *J Immunol* 1998, **160**:3363-3373.