In silico Immunogenicity Assessment for Sequences Containing Unnatural Amino Acids:

A method using existing in silico algorithm infrastructure and a vision for future enhancements

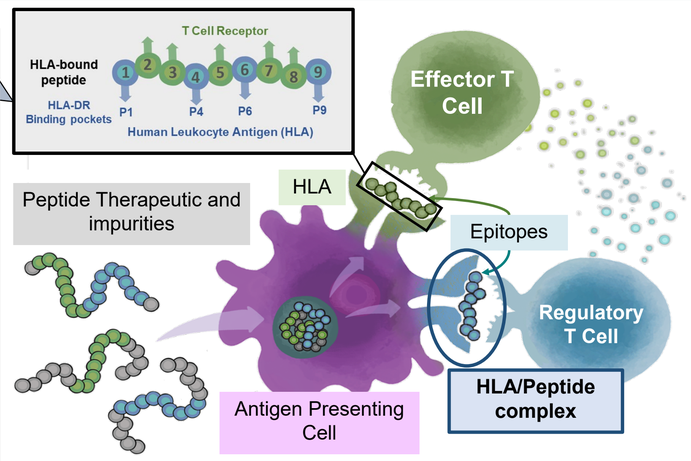
Aimee Mattei MS, Andres Gutierrez PhD, William Martin, Frances Terry MPH, Brian J Roberts PhD, Amy S. Rosenberg, Anne S. De Groot MD
EpiVax, Inc. Providence, RI, United States

BACKGROUND INFROMATION

In silico prediction of T cell epitopes within a peptide drug candidate serves as an important 1st step for assessing immunogenicity.

T cell epitopes bind HLA by a well-characterized interaction of amino acid side chains and pockets in the HLA-DR molecule binding groove.

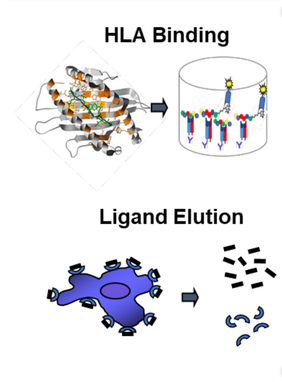
Immunoinformatics tools, such as **EpiMatrix**, have been developed to screen natural amino acid sequences for peptides that will bind HLA.



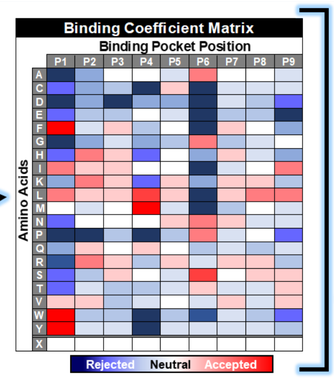
In silico assessment of immunogenic potential allows for risk-based selection of best candidate peptides in further confirmatory in vitro, ex vivo and in vivo assays, thereby reducing the overall cost of immunogenicity evaluation.

Immunogenicity Scores are calculated from *in silico* algorithms trained on curated *in vitro* data for natural AA peptide sequences

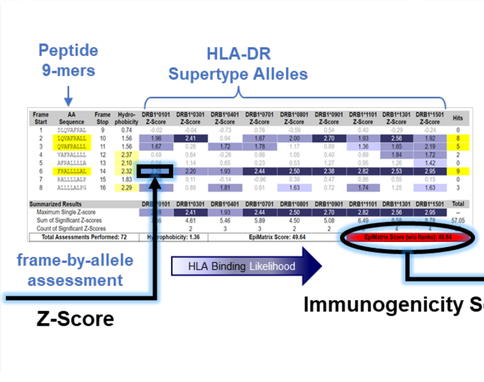
In vitro Data Curation and Frequency Analysis



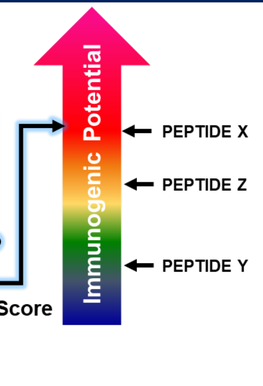
Training the Algorithm (Allele-Specific Matrix)



Predicting HLA Ligands (EpiMatrix Detail Report)



Generating a Score (Immunogenicity Scale)



HLA binding properties of peptides containing unnatural amino acids (UAA) are not accurately estimated by most algorithms, to date.

UAA are often incorporated into peptide therapeutics to improve drug properties and commonly occur in synthetic peptide-related impurities. Both scenarios warrant the need for enhanced predictive algorithms.

OVERVIEW & ROAD MAP

Process to expand existing in silico immunogenicity prediction tools to handle sequences containing unnatural amino acids

Phase 1 — Substitution Method with EXISTING infrastructure

Phase 2 — Apply '**correction factors**' to common UAA
- a. Expert Review
- b. Validation in vitro
- c. Validation ex vivo

Phase 3 — Common UAAs programmed for **direct scoring.***

*requires extensive in vitro data to train the algorithm

1. Natural amino acid or placeholder substitutions must be applied to score peptides that include UAA with existing epitope mapping tools. (Phase 1)
2. 'Correction factors' can be applied to the scores of natural amino acid substitutions to more accurately predict HLA binding for UAA-containing sequences. (Phase 2)
   a) D-amino acids and side chain modifications that introduce 'bulk' are expected to negatively impact the HLA binding likelihood and can be modeled by introducing a deduction to the score of the closest matching natural L-amino acid. (Phase 2a)
   b) These 'correction factors' will be further refined based on in vitro validation data. (Phase 2b,c)

PHASE 2: Apply "Correction Factors" to common UAA

a) Expert Review
Review of UAA side chain structure compared to closest matching natural AA and apply score deductions. (i.e. minimal, moderate, or significant).

b) Validation *in vitro*
HLA binding assays to compare known ligand sequences modified with selected UAA in HLA binding positions. See Below for example with D-AA.

c) Validation *ex vivo*
IVIP T cell assays to assess the impact of selected UAA on T cell recognition and immunogenic potentials.

Preliminary *in vitro* data for D-amino acid 'correction factors'

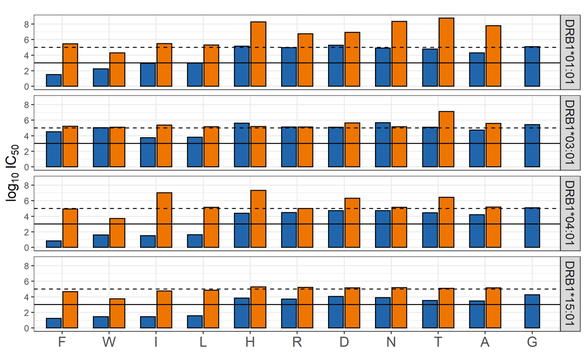
HLA binding studies based on known ligands (such as PADRE) which have been modified to contain UAAs in HLA anchoring positions (1,4,6,9) can be used to estimate the binding potentials of commonly encountered UAA and "correct" *in silico* estimates of binding based on their naturally-occurring counterparts.

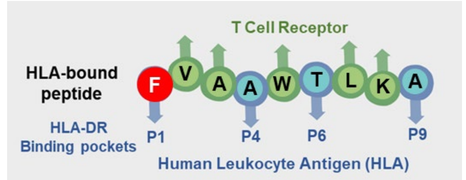
EpiMatrix Detail Report - PADRE PEPTIDE

Frame Start	AA Seqeunce	Frame Stop	DRB1* 0101	DRB1* 0301	DRB1* 0401	DRB1* 0701	DRB1* 0801	DRB1* 0901	DRB1* 1101	DRB1* 1301	DRB1* 1501	Hits
1	AKFVAAWTL	9										0
2	KFVAAWTLK	10										0
3	FVAAWTLKA	11										9
4	VAAWTLKAA	12										0
5	AAWTLKAAA	13										0

EpiMatrix Score : 17.12

Top 10% of random peptides (not significant, near miss) | Top 5% of random peptides (significant hit) | Top 1% of random peptides (significant hit, highly likely)





D-amino acids disrupt HLA binding compared to their corresponding L-amino acid isomer when substituted in HLA binding position 1 of a known promiscuous HLA-DR binding peptide (PADRE).

Similar studies are underway assessing the impact of D-AA in other positions. These studies will enable calculation of correction factors for adjusting in silico HLA binding estimates for D-AA, and to further adjust the binding predictions for D-AA by allele and amino acid group.

PHASE 1 : Substitution Method with Existing *in silico* Algorithms

Three Steps to Select a Best Proxy Substitution for the Unnatural Amino Acid

1. Neutral Placeholder
Replace the UAA with a neutral placeholder X. In EpiMatrix, amino acid "X" has a coefficient of 0 and is assumed to neither promote nor detract from binding.

2. Replacement Analysis
Replace the UAA with each of the 20 natural L-amino acids to establish the extent to which variation at this position can have an impact on the binding potential of the input peptide and identify which, if any, substitutes are likely to promote or significantly detract from HLA binding potential.

3. Structural Proxy
Review the structural/ chemical properties of the UAA side chain and, if applicable, replace with the closest matching natural L-amino acid.

Example : Semaglutide API and D-Amino Acid Impurities

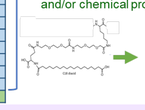
Selecting suitable replacements for the UAA in Semaglutide API

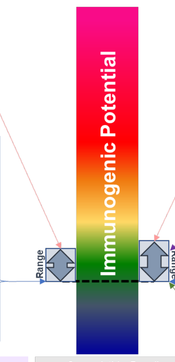
Aminoisobutyric Acid (Aib) | **K(OEG-OEG-γGlu-C18diacid)**

SEMAGLUTIDE API SEQUENCE
HAibEGTFTSDVSSYLEGQAAK(OEG-OEG-γGlu-C18diacid)EFIAWLVRGRG

1 Neutral placeholder: Replace UAA (Aib) with X → HXEGTFTSDVSSYLEGQAAZEFIAWLVRGRG
2 Replacement Analysis: Replace X with all 20 natural L-amino acids
Highest Score = most conducive to binding
Lowest Score = least conducive to binding
3 Structural Proxy: Choose best-matching natural L-amino acid based on structural and/or chemical properties

1 Neutral placeholder: Replace UAA (K(OEG-OEG-γGlu-C18 diacid) with X → HAEGTFTSDVSSYLEGQAAXEFIAWLVRGRG
2 Replacement Analysis: Replace X with all 20 natural L-amino acids
Highest Score = most conducive to binding
Lowest Score = least conducive to binding
3 Structural Proxy: Choose best-matching natural L-amino acid based on structural and/or chemical properties





Aminoisobutyric acid (Aib)
- Narrow range of scores → little impact on HLA binding potential
- Replace with structural proxy, Ala.

K(OEG-OEG-γGlu-C18diacid)
- Narrow range of scores → little impact on HLA binding potential
- Expected to disrupt potential HLA binding, replace with low affinity placeholder, Z.

EpiMatrix analysis for Semaglutide API

EpiMatrix Detail Report - SEMAGLUTIDE

Frame Start	AA Seqeunce	Frame Stop	HLA-DR Supertype Alleles DRB1* 0101	DRB1* 0301	DRB1* 0401	DRB1* 0701	DRB1* 0801	DRB1* 0901	DRB1* 1101	DRB1* 1301	DRB1* 1501	Hits
7	HAEGTFTSD	15										0
8	AEGTFTSDV	16										0
9	EGTFTSDVS	17										0
10	GTFTSDVSS	18										0
11	TFTSDVSSY	19										0
12	FTSDVSSYL	20										6
13	TSDVSSYLE	21										0
14	SDVSSYLEG	22										0
15	DVSSYLEGQ	23										0
16	VSSYLEGQA	24										0
17	SSYLEGQAA	25										0
18	SYLEGQAAZ	26										0
19	YLEGQAAZE	27										0
20	LEGQAAZEF	28										0
21	EGQAAZEFI	29										0
22	GQAAZEFIA	30										0
23	QAAZEFIAW	31										0
24	AAZEFIAWL	32										0
25	AZEFIAWLV	33										0
26	ZEFIAWLVR	34										0
27	EFIAWLVRG	35										0
28	FIAWLVRGR	36										2
29	IAWLVRGRG	37										2

EpiMatrix Score : -0.25

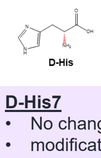
Top 10% of random peptides (not significant, near miss) | Top 5% of random peptides (significant hit) | Top 1% of random peptides (significant hit, highly likely)

*note: semaglutide peptide numbering is relative to hGLP-1 (7-37)

Semaglutide D-amino acid impurity analysis examples

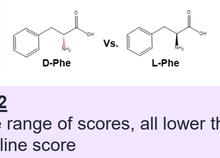
D-His in position 7*

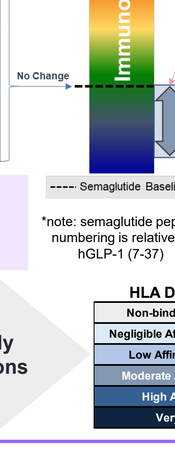
1 Neutral placeholder: Replace UAA (D-His) with X → hAEGTFTSDVSSYLEGQAAZEFIAWLVRGRG / XAEGTFTSDVSSYLEGQAAZEFIAWLVRGRG
2 Replacement Analysis: Replace X with all 20 natural L-amino acids
Highest Score = most conducive to binding
Lowest Score = least conducive to binding
3 Structural Proxy: Choose best-matching natural L-amino acid based on structural and/or chemical properties



D-Phe in position 12*

1 Neutral placeholder: Replace UAA (D-Phe) with X → HAEGTfTSDVSSYLEGQAAZEFIAWLVRGRG / HAEGTXTSDVSSYLEGQAAZEFIAWLVRGRG
2 Replacement Analysis: Replace X with all 20 natural L-amino acids
Highest Score = most conducive to binding
Lowest Score = least conducive to binding
3 Structural Proxy: Choose best-matching natural L-amino acid based on structural and/or chemical properties





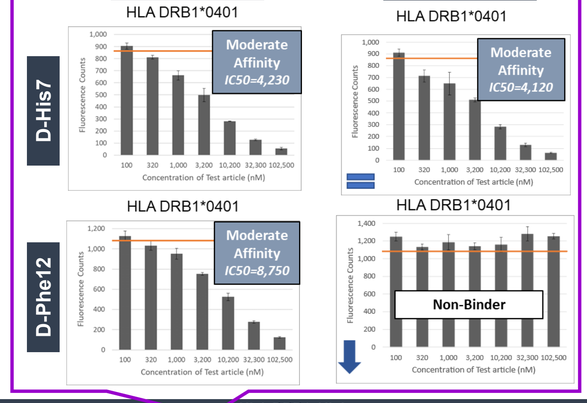
*note: semaglutide peptide numbering is relative to hGLP-1 (7-37)

D-His7
- No change in score
- modification will not impact HLA binding

D-Phe12
- Wide range of scores, all lower than baseline score
- modification will disrupt HLA binding

In Vitro HLA Binding Study Confirms in Silico Predictions

HLA DR Binding Affinity Cutoffs
- Non-binder (No dose-dependent inhibition)
- Negligible Affinity (100,000nM < IC_{50}<1,000,000nM)
- Low Affinity (10,000nM < IC_{50} < 100,000nM)
- Moderate Affinity (1,000nM < IC_{50} < 10,000nM)
- High Affinity (100nM < IC_{50} < 1,000nM)
- Very High Affinity (IC_{50} < 100nM)



		HLA-DRB1 Alleles *0101	*0301	*0401	*0701	*0901	*1101	*1501
D-His7	BASELINE: 00_SEM_API(7-23)	Low	Low	Moderate	Low	Low	Non-Binder	Negligible
D-His7	IMPURITY: D-HIS7_SEM(7-23)	Negligible ↓	Low =	Moderate =	Low =	Low =	Non-Binder =	Negligible =
D-Phe12	BASELINE: 00_SEM_API(9-23)	Non-Binder	Low	Moderate	Low	Low	Non-Binder	Negligible
D-Phe12	IMPURITY: D-PHE12_SEM(9-23)	Non-Binder =	Negligible ↓	Non-Binder ↓	Negligible ↓	Negligible ↓	Non-Binder =	Negligible =

TAKE HOME MESSAGE

- *In silico* risk assessment of peptides and their related impurities is an important first step to understanding the immunogenic potential of a given therapeutic, but *in silico* immunogenicity prediction algorithms are limited to natural amino acid sequences.
- A three-phased approach for the eventual incorporation of common unnatural amino acids into immunoinformatic toolkits includes: first, a substitution-based method enabling *in silico* immunogenicity risk assessment for sequences containing unnatural amino acids, and second, the use of in vitro HLA binding and ex vivo T cell assays in the development of 'correction factors' that can be applied to *in silico* 'scores' for common unnatural amino acids, for more accuracy in predictions.

FUNDING & REFERENCES

Some of the data presented was funded in part by FDA Contract # 75F40120C00157.

Mattei AE, Gutierrez AH, Martin WD,Terry FE, Roberts BJ, Rosenberg AS and De Groot AS (2022), Front. Drug. Discov. 2:952326. doi: 10.3389/fddsv.2022.952326