Predicting Immunogenicity of Peptide Drugs and their Impurities using in Silico tools: Taspoglutide Case Study

Aimee Mattei MS[1], Frances Terry MPH[1], Brian J Roberts PhD[1], William Martin[1], and Anne S De Groot MD[1,2]

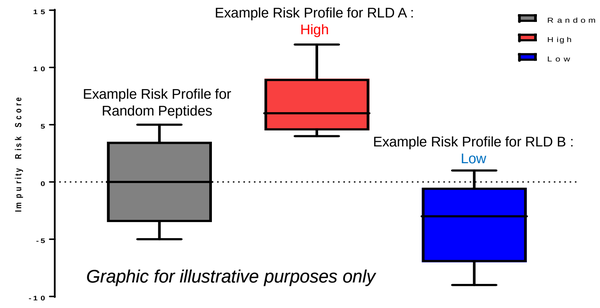
[1]EpiVax, Inc.; [2]Professor (Research) and Director, Institute for Immunology and Informatics, University of Rhode Island, USA

Abstract

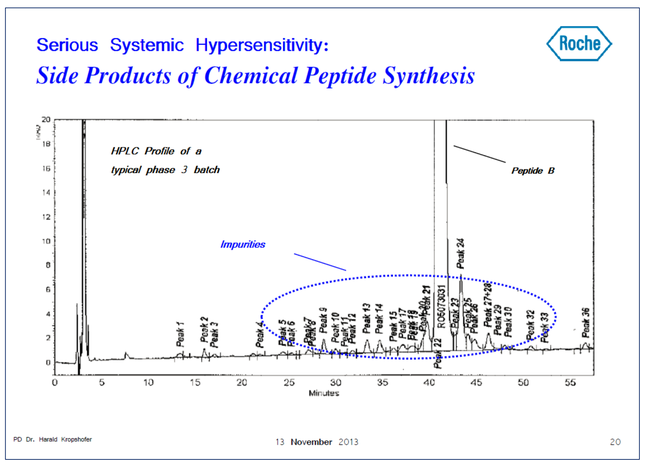
- The peptide drug market is expected to generate $50B in revenue by 2024 but the FDA is concerned about the number of impurities that may be introduced in the synthetic process.
- The peptide manufacturing process can result in synthesis-related impurities that can introduce immunogenic epitopes within the amino acid sequence of the peptide, resulting in an unexpected and undesired immune response against the drug.
- EpiMatrix can be used to screen both the drug API sequence and its known peptide-related impurities for the presence of putative T cell epitope content.
- When peptide-related impurities are unknown, the "What if Machine" (WhIM) can perform theoretical changes to the natural amino acid sequence of the drug substance and measure their impact on the putative epitope content of the peptide.
- Here we present a retrospective case study of Taspoglutide – a GLP-1 agonist that was under investigation for the treatment of type 2 diabetes, but development was halted during phase III clinical trials due to serious hypersensitivity reactions and GI intolerance.[1]
- Using the WhIM algorithm, we have evaluated all possible amino acid duplication impurities for the presence of new T cell epitopes at both a population level and an individualized level.
- We identified five impurities that could be contributing to the observed hypersensitivity to Taspoglutide.

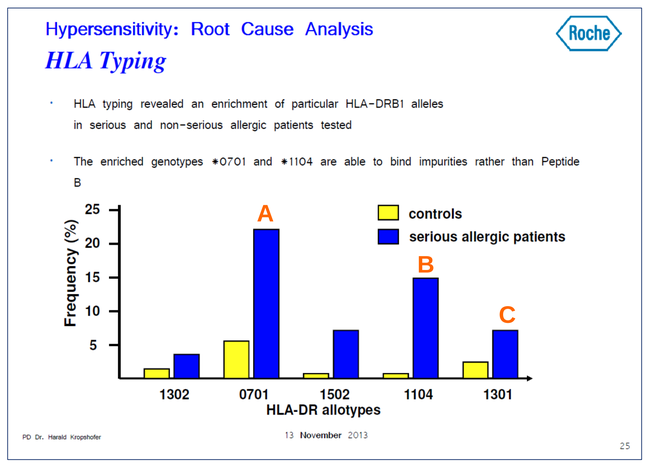
What-if-Machine (WhIM)

- Mimics the process of synthesizing polypeptides and records theoretical product impurities created through known failures in the synthesis process.
- Each identified impurity is scored for putative T cell epitope content (EpiMatrix) and cross conservation with the human proteome (JanusMatrix).
- Impurities are weighted based on assumed probability of occurrence.



Taspoglutide and Observed Hypersensitivity





- It is suspected that the cause of the observed hypersensitivity is due to the presence of amino acid duplication synthesis side product(s) which gives rise to novel T cell epitopes.
- HLA typing in allergic patients shows an enrichment of five particular HLA DRB1 alleles. Two of these alleles (DRB1*0701 and DRB1*1104) were shown to be able to bind the impurity rather than the drug.[2]

Taspoglutide What-if-Machine Analysis Results

- It is suspected that an amino acid duplication impurity that creates a neo-epitope was the cause of the observed hypersensitivity to Taspoglutide. (below)
- Using the WhIM algorithm, all duplication impurities were analyzed with EpiMatrix and JanusMatrix for immunogenic potential. (right)

EpiMatrix (EMX) Hits:
- aa duplication impurities with more EMX hits than baseline (9) are in **red** → **increase** in overall putative T cell epitope content
- aa duplication impurities with fewer EMX hits than baseline (9) in **green** → **decrease** in overall putative T cell epitope content

JMX HMLGY Score:
- JanusMatrix Human Homology Score
- **Scores > 3** are considered significant for potential homology induced tolerance

Input Sequence	Peptide Sequence	EpiMatrix Hits	EpiMatrix Score	EpiBars? (#)	Number of HUMAN Matches	Janus HMLGY Score
00_HGLP-1	HAEGTFTSDVSSYLEGQAAKEFIAWLVKAR	9	0.86	Yes (1)	9	1.67
01_ENDO-HIS7_HGLP-1	HHAEGTFTSDVSSYLEGQAAKEFIAWLVKAR	9	0.04	Yes (1)	9	1.67
02_ENDO-ALA8_HGLP-1	HAAEGTFTSDVSSYLEGQAAKEFIAWLVKAR	9	0.04	Yes (1)	9	1.67
03_ENDO-GLU9_HGLP-1	HAEEGTFTSDVSSYLEGQAAKEFIAWLVKAR	9	0.04	Yes (1)	9	1.67
04_ENDO-GLY10_HGLP-1	HAEGGTFTSDVSSYLEGQAAKEFIAWLVKAR	9	0.04	Yes (1)	9	1.67
05_ENDO-THR11_HGLP-1	HAEGTTFTSDVSSYLEGQAAKEFIAWLVKAR	9	0.04	Yes (1)	9	1.67
06_ENDO-PHE12_HGLP-1	HAEGTFFTSDVSSYLEGQAAKEFIAWLVKAR	10	1.72	Yes (1)	9	1.50
07_ENDO-THR13_HGLP-1	HAEGTFTTSDVSSYLEGQAAKEFIAWLVKAR	5	-9.91	No	6	1.40
08_ENDO-SER14_HGLP-1	HAEGTFTSSDVSSYLEGQAAKEFIAWLVKAR	6	-7.75	No	7	1.33
09_ENDO-ASP15_HGLP-1	HAEGTFTSDDVSSYLEGQAAKEFIAWLVKAR	5	-9.22	No	7	1.60
10_ENDO-VAL16_HGLP-1	HAEGTFTSDVVSSYLEGQAAKEFIAWLVKAR	8	-3.12	No	28	**6.63**
11_ENDO-SER17_HGLP-1	HAEGTFTSDVSSSYLEGQAAKEFIAWLVKAR	8	-2.19	Yes (1)	27	**8.38**
12_ENDO-TYR19_HGLP-1	HAEGTFTSDVSSYYLEGQAAKEFIAWLVKAR	7	-4.55	No	8	1.43
13_ENDO-LEU20_HGLP-1	HAEGTFTSDVSSYLLEGQAAKEFIAWLVKAR	8	-1.66	Yes (1)	6	1.50
14_ENDO-GLU21_HGLP-1	HAEGTFTSDVSSYLEEGQAAKEFIAWLVKAR	8	-1.66	Yes (1)	6	1.50
15_ENDO-GLY22_HGLP-1	HAEGTFTSDVSSYLEGGQAAKEFIAWLVKAR	8	-1.66	Yes (1)	6	1.50
16_ENDO-GLN23_HGLP-1	HAEGTFTSDVSSYLEGQQAAKEFIAWLVKAR	10	2.23	Yes (1)	6	1.20
17_ENDO-ALA24_HGLP-1	HAEGTFTSDVSSYLEGQAAAKEFIAWLVKAR	10	2.15	Yes (1)	10	1.60
18_ENDO-LYS26_HGLP-1	HAEGTFTSDVSSYLEGQAAKKEFIAWLVKAR	12	5.25	Yes (1)	11	1.58
19_ENDO-GLU27_HGLP-1	HAEGTFTSDVSSYLEGQAAKEEFIAWLVKAR	9	0.04	Yes (1)	9	1.67
20_ENDO-PHE28_HGLP-1	HAEGTFTSDVSSYLEGQAAKEFFIAWLVKAR	12	5.72	Yes (1)	9	1.25
21_ENDO-ILE29_HGLP-1	HAEGTFTSDVSSYLEGQAAKEFIIAWLVKAR	11	4.34	Yes (2)	7	1.09
22_ENDO-ALA30_HGLP-1	HAEGTFTSDVSSYLEGQAAKEFIAAWLVKAR	10	2.32	Yes (2)	6	1.10
23_ENDO-TRP31_HGLP-1	HAEGTFTSDVSSYLEGQAAKEFIAWWLVKAR	7	-3.62	Yes (1)	6	1.57
24_ENDO-LEU32_HGLP-1	HAEGTFTSDVSSYLEGQAAKEFIAWLLVKAR	8	-1.60	Yes (1)	8	1.63
25_ENDO-VAL33_HGLP-1	HAEGTFTSDVSSYLEGQAAKEFIAWLVVKAR	9	1.09	Yes (1)	8	1.56
26_ENDO-LYS34_HGLP-1	HAEGTFTSDVSSYLEGQAAKEFIAWLVKKAR	12	6.35	Yes (2)	9	1.33
27_ENDO-ALA35_HGLP-1	HAEGTFTSDVSSYLEGQAAKEFIAWLVKAAR	13	8.04	Yes (2)	9	1.15
28_ENDO-ARG36_HGLP-1	HAEGTFTSDVSSYLEGQAAKEFIAWLVKARR	13	7.78	Yes (2)	9	1.15

EpiMatrix Cluster Detail Report

File: TASPOGLUTIDE Sequence: 26_ENDO-LYS34_HGLP-1 Cluster: 7

Frame Start	AA Sequence	Frame Stop	Hydro-phobicity	DRB1*0101 Z-Score	DRB1*0301 Z-Score	DRB1*0401 Z-Score	DRB1*0701 Z-Score	DRB1*0801 Z-Score	DRB1*1101 Z-Score	DRB1*1301 Z-Score	DRB1*1501 Z-Score	Hits	Different from baseline (Y/N)	Neo-epitopes
7	HAEGTFTSD	15	-0.91	-1.99	-1.54	-1.97	-1.57	-1.08	-2.78	-1.96	-1.97	0	N	0
8	AEGTFTSDV	16	-0.09	-0.09	-0.87	0.00	0.58	-0.91	-1.17	-0.13	-0.12	0	N	0
9	EGTFTSDVS	17	-0.38	-0.21	-1.72	0.18	-1.31	-1.13	-0.48	-0.99	-0.28	0	N	0
10	GTFTSDVSS	18	-0.08	-0.39	-0.58	0.57	-0.48	-0.16	-0.94	0.24	0.31	0	N	0
11	TFTSDVSSY	19	-0.18	-0.64	0.20	-0.54	-0.54	-1.85	-0.99	-0.23	-1.58	0	N	0
12	FTSDVSSYL	20	0.32	2.29	2.83	2.55	2.27	1.52	0.92	1.60	1.75	5	N	0
13	TSDVSSYLE	21	-0.38	-0.89	-0.84	-0.61	-0.36	-0.16	-0.70	-0.75	-0.61	0	N	0
14	SDVSSYLEG	22	-0.34	-0.31	-0.55	-0.75	-0.85	-1.27	-0.77	-0.72	-0.09	0	N	0
15	DVSSYLEGQ	23	-0.64	-0.70	0.49	-0.12	-0.06	0.05	0.55	-0.31	-0.29	0	N	0
16	VSSYLEGQA	24	-0.06	-0.84	0.02	-0.67	0.37	0.47	0.12	0.25	0.86	0	N	0
17	SSYLEGQAA	25	-0.32	0.75	-0.20	0.31	-0.12	-0.30	-0.08	-1.16	0.35	0	N	0
18	SYLEGQAAK	26	-0.67	-0.17	-0.49	-0.15	-1.15	-1.12	-0.15	-0.48	-1.90	0	N	0
19	YLEGQAAKE	27	-0.97	0.97	1.33	1.24	0.42	1.70	0.96	1.07	0.31	1	N	0
20	LEGQAAKEF	28	-0.51	0.91	0.62	0.90	1.26	0.92	0.97	1.55	0.41	0	N	0
21	EGQAAKEFI	29	-0.43	-0.34	0.76	-0.95	-0.44	0.26	-0.28	0.27	-0.15	0	N	0
22	GQAAKEFIA	30	0.16	-0.06	0.19	-0.66	0.18	-0.39	0.26	0.11	0.69	0	N	0
23	QAAKEFIAW	31	0.1	-1.99	-1.47	-2.16	-0.04	-0.84	-1.41	-1.02	-2.39	0	N	0
24	AAKEFIAWL	32	0.91	-0.49	-0.46	-0.33	-0.15	-1.61	-0.69	-0.45	-1.50	0	N	0
25	AKEFIAWLV	33	1.18	0.70	-0.70	0.26	0.77	-0.70	0.70	-0.01	1.43	0	N	0
26	KEFIAWLVK	34	0.54	-0.57	-1.20	0.27	0.05	-1.18	-0.02	-0.68	0.28	0	N	0
27	EFIAWLVKK	35	0.54	-0.58	0.13	-0.08	-0.73	0.35	-0.33	0.11	0.02	0	Y	0
28	FIAWLVKKA	36	1.13	1.33	0.61	2.45	2.16	0.99	1.81	1.80	0.68	4	Y	4
29	IAWLVKKAR	37	0.32	0.83	1.26	-0.02	0.08	2.01	1.48	1.68	0.92	2	Y	2

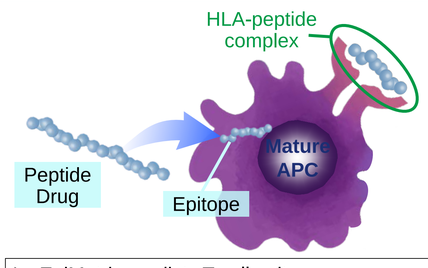
Summarized Results	DRB1*0101	DRB1*0301	DRB1*0401	DRB1*0701	DRB1*0801	DRB1*1101	DRB1*1301	DRB1*1501	Total	Total
Maximum Single Z-score	2.29	2.83	2.55	2.27	2.01	1.81	1.80	1.75	--	6
Sum of Significant Z-scores	2.29	2.83	5.00	4.43	3.71	1.81	3.48	1.75	25.30	
Count of Significant Z-Scores	1	1	2	2	2	1	2	1	12	
Total Assessments Performed: 184	Hydrophobicity: -0.28		EpiMatrix Score: 6.35			EpiMatrix Score (w/o flanks): 6.35				
Scores Adjusted for Tregitope:	--		EpiMatrix Score: 6.35			EpiMatrix Score (w/o flanks): 6.35				

Neo-epitopes for:
- A → DRB1*0701
- B → DRB1*1101
- C → DRB1*1301

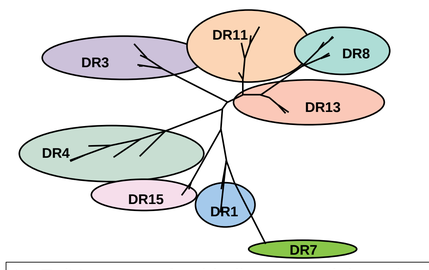
Disclaimer:
This exploratory analysis considered theoretical impurities of Taspoglutide resulting from amino acid duplication only. Other peptide-related impurities that may be present in the drug product could also contribute to the observed hypersensitivity

In Silico Evaluation of Immunogenicity

Searching for T Cell Epitopes with EpiMatrix



- EpiMatrix predicts T cell epitopes
- HLA binding is a prerequisite for immunogenicity



- EpiVax tests for binding potential to the most common HLA molecules within each of the "supertypes"[3] shown above.
- These are representative of >95% of human populations worldwide.[4]

EpiMatrix Cluster Detail Report

File: TASPOGLUTIDE Sequence: 00_HGLP-1 Cluster: 7

Frame Start	AA Sequence	Frame Stop	Hydro-phobicity	DRB1*0101 Z-Score	DRB1*0301 Z-Score	DRB1*0401 Z-Score	DRB1*0701 Z-Score	DRB1*0801 Z-Score	DRB1*1101 Z-Score	DRB1*1301 Z-Score	DRB1*1501 Z-Score	Hits
7	HAEGTFTSD	15	-0.91	-1.99	-1.54	-1.97	-1.57	-1.08	-2.78	-1.96	-1.97	0
8	AEGTFTSDV	16	-0.09	-0.09	-0.87	0.00	0.58	-0.91	-1.17	-0.13	-0.12	0
9	EGTFTSDVS	17	-0.38	-0.21	-1.72	0.18	-1.31	-1.13	-0.48	-0.99	-0.28	0
10	GTFTSDVSS	18	-0.08	-0.39	-0.58	0.57	-0.48	-0.16	-0.94	0.24	0.31	0
11	TFTSDVSSY	19	-0.18	-0.64	0.20	-0.54	-0.54	-1.85	-0.99	-0.23	-1.58	0
12	FTSDVSSYL	20	0.32	2.29	2.83	2.55	2.27	1.52	0.92	1.60	1.75	5
13	TSDVSSYLE	21	-0.38	-0.89	-0.84	-0.61	-0.36	-0.16	-0.70	-0.75	-0.61	0
14	SDVSSYLEG	22	-0.34	-0.31	-0.55	-0.75	-0.85	-1.27	-0.77	-0.72	-0.09	0
15	DVSSYLEGQ	23	-0.64	-0.70	0.49	-0.12	-0.06	0.05	0.55	-0.31	-0.29	0
16	VSSYLEGQA	24	-0.06	-0.84	0.02	-0.67	0.37	0.47	0.12	0.25	0.86	0
17	SSYLEGQAA	25	-0.32	0.75	-0.20	0.31	-0.12	-0.30	-0.08	-1.16	0.35	0
18	SYLEGQAAK	26	-0.67	-0.17	-0.49	-0.15	-1.15	-1.12	-0.15	-0.48	-1.90	0
19	YLEGQAAKE	27	-0.97	0.97	1.33	1.24	0.42	1.70	0.96	1.07	0.31	1
20	LEGQAAKEF	28	-0.51	0.91	0.62	0.90	1.26	0.92	0.97	1.55	0.41	0
21	EGQAAKEFI	29	-0.43	-0.34	0.76	-0.95	-0.44	0.26	-0.28	0.27	-0.15	0
22	GQAAKEFIA	30	0.16	-0.06	0.19	-0.66	0.18	-0.39	0.26	0.11	0.69	0
23	QAAKEFIAW	31	0.1	-1.99	-1.47	-2.16	-0.04	-0.84	-1.41	-1.02	-2.39	0
24	AAKEFIAWL	32	0.91	-0.49	-0.46	-0.33	-0.15	-1.61	-0.69	-0.45	-1.50	0
25	AKEFIAWLV	33	1.18	0.70	-0.70	0.26	0.77	-0.70	0.70	-0.01	1.43	0
26	KEFIAWLVK	34	0.54	-0.57	-1.20	0.27	0.05	-1.18	-0.02	-0.68	0.28	0
27	EFIAWLVKA	35	1.18	-0.25	0.51	0.06	-0.36	0.17	-0.17	0.48	0.34	0
28	FIAWLVKAR	36	1.07	0.98	0.83	1.85	1.72	1.26	1.17	2.02	0.34	3

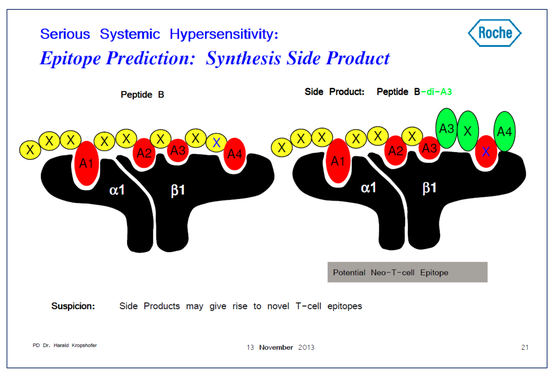
Summarized Results	DRB1*0101	DRB1*0301	DRB1*0401	DRB1*0701	DRB1*0801	DRB1*1101	DRB1*1301	DRB1*1501	Total
Maximum Single Z-score	2.29	2.83	2.55	2.27	1.70	1.17	2.02	1.75	--
Sum of Significant Z-scores	2.29	2.83	4.40	3.99	1.70	0.00	2.02	1.75	18.98
Count of Significant Z-Scores	1	1	2	2	1	0	1	1	9
Total Assessments Performed: 176	Hydrophobicity: -0.16		EpiMatrix Score: 0.86			EpiMatrix Score (w/o flanks): 0.86			
Scores Adjusted for Tregitope:	--		EpiMatrix Score: 0.86			EpiMatrix Score (w/o flanks): 0.86			

3 frames contain putative T cell epitopes (Z-scores ≥ 1.64, medium and dark blue shading)

EpiBar = promiscuous binding motif

Taspoglutide has a total of 9 EpiMatrix Hits

Neutral EpiMatrix Score (scores above 10 indicate significant immunogenic potential)



28 possible aa duplication impurities

10 with more T cell epitopes compared to Taspoglutide

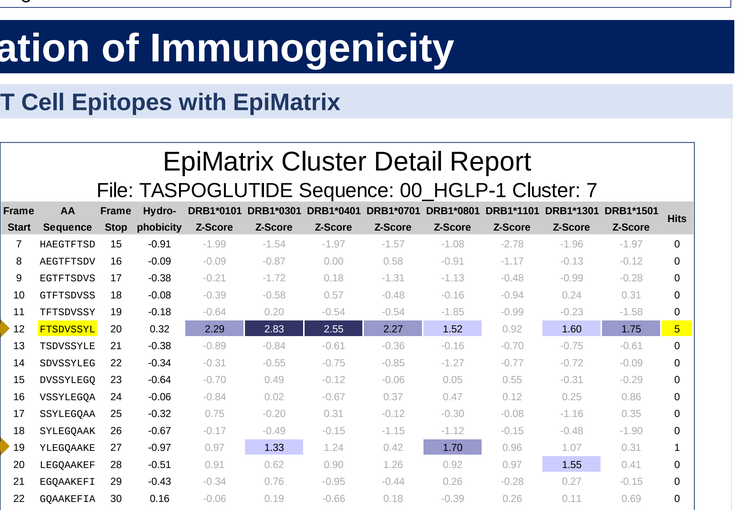
5 create neo-epitopes for the HLA-DR7 and/or DR11 families

One or more could be contributing to **hypersensitivity** in subjects with DRB1*0701 and/or DRB1*1104

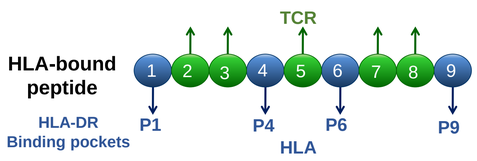
Neo-epitopes introduced as a result of the impurity are outlined in the EpiMatrix Detail Report above in red.
- Modifications that **create** an HLA ligand that is not present in the API
- Modifications that preserve a ligand predicted in the API but have a **change at the TCR face**.

Peptide-Related Impurities and Neo-epitopes

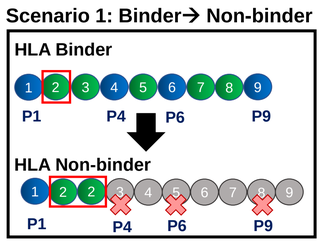
1. Peptides are synthesized from C-term to N-term.
2. Addition of each amino acid is termed a "cycle" (Cn).
3. During any given cycle an impurity can be introduced (right). Impurities can create **neo-epitopes** (below).



Example Impurity – AA Duplication



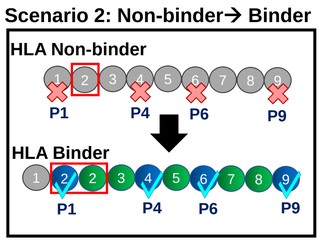
Scenario 1: Binder→ Non-binder



results in a peptide that will no longer bind HLA by shifting subsequent amino acids out of phase

Low-Risk Impurity*

Scenario 2: Non-binder→ Binder



results in a peptide that will now bind HLA (*neo-epitope*) by shifting subsequent amino acids into phase

Potentially Immunogenic Impurity*

*Based on T cell epitope content alone

Conclusions

- **It is important to assess the potential immunogenicity of not only peptide drug candidates, but also their synthesis-related impurities in early stages of drug development.**
- **In silico tools such as EpiMatrix and JanusMatrix can provide a quick and cost-effective method to screen peptides for immunogenicity.**
- **When impurities are unknown, the What-if-Machine can quickly screen all plausible peptide-related impurity sequences and identify potentially immunogenic impurities.**
- **A retrospective WhIM analysis of Taspoglutide revealed five potential impurities that could contribute to the observed hypersensitivity and consequently, the halting of the development of Taspoglutide.**

References

1. Rosenstock et al. The Fate of Taspoglutide, a weekly GLP-1 Receptor Agonist, Versus Twice-Daily Exenatide for Type 2 Diabetes. The T-emerge 2 trial. Diabetes Care. 2013; 36:498-504
2. Thank you to Dr. Harald Kropshofer for providing the Taspoglutide peptide impurity and HLA typing graphs.
3. Lund et al. Definition of Supertypes for HLA Molecules Using Clustering of Specificity Matrices. Immunogenetics. 2004; 55(12):797–810.
4. Southwood et al. Several Common HLA-DR Types Share Largely Overlapping Peptide Binding Repertoires. J Immunol. 1998; 160(7):3363–73.