In Silico Immunogenicity Profiling Based On Drug / Pathogen Analogy

Author(s): Miriam Fuhlendorf[1], Anette Karle[1], Martin Jockel[2], Guillaume Roellinger[2], Pierre Dönnes[3], Michael Gutknecht[1]

Affiliations: [1]Novartis Pharma AG, Immunogenicity & Mechanistic Immunology, Basel, Switzerland; [2]Novartis Pharma AG, Bioinformatics & Biotherapeutic Modeling, Basel, Switzerland; [3]SciCross, Sweden

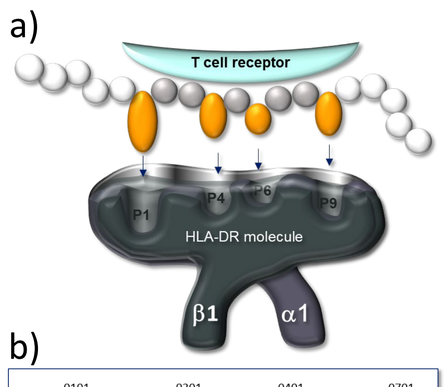
Abstract

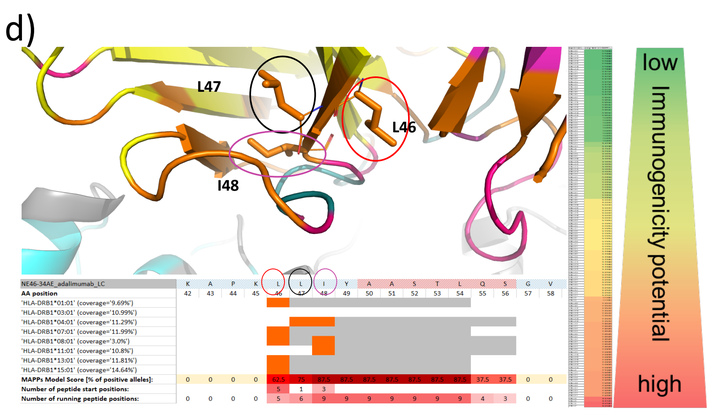
Immunogenicity potential assessment should be started as early as possible in the biotherapeutic development process to inform necessary de-immunization approaches and to avoid resources spending on candidates with a high inherent IG potential that can't be resolved in a reasonable way. Oftentimes, this is only possible using in silico tools, since in early drug development, high-quality candidate material is not available in the quantities necessary for most in vitro assays. Additionally, high cost and long timelines of in vitro assays are also factors that can be prohibitive for pharma and biotech companies alike. Most in silico tools used in biotherapeutic development are predicting peptide binding to HLA class II molecules (e.g. NetMHCIIpan), frequently with the option to apply a weighting matrix, based on the hypothesis that self-peptides and germline sequences have a lower IG potential. Based on our experience during recent root cause analysis of adverse events, we started to explore additional options to improve this weighting matrix. We could show that biotherapeutic sequences can bear analogues to pathogen sequences, which may result in a high number of memory T cells that are cross-specific to the biotherapeutic, as well as a high prevalence of pre-existing anti-drug antibodies. So, for a comprehensive in silico IG analysis, not only germline whitelisting and self-peptide filtering should be applied but also a pathogen database lookup to identify analogous peptides that could lead to the development of a strong IG response in a larger proportion of the patient population.

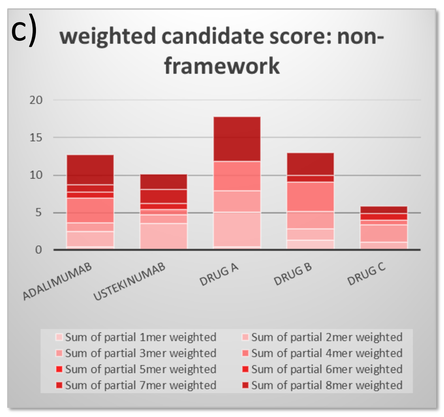
In Silico Immunogenicity Profiling At Novartis

Prediction of HLA class II binding based on a Position Specific Scoring Model (PSSM)

- ❑ The PSSM algorithm called "iSHAPe" (*in silico* HLA Aggretope Prediction) was trained on superior in-house MAPPs assay data, consisting of over 20 million heterozygous peptides and over 100K homozygous peptides.
- ❑ The "iSHAPe" algorithm calculates the likelihood that a specific peptide can bind to a particular HLA molecule (a,b).
- ❑ iSHAPe is used for immunogenicity profiling of large candidate sets using a weighting matrix based on hotspot and CDR overlap ranking (c).
- ❑ Protein design is supported via de-immunization of sequence hotspots (d).





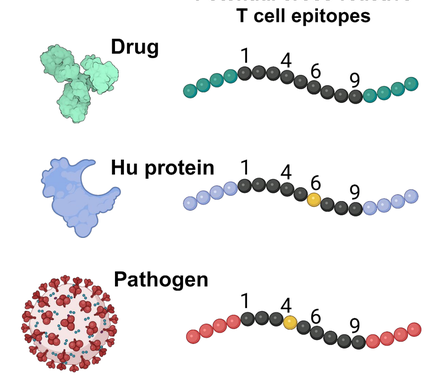


Correlation Of Clinical IG And Cross-reactive T Cell Epitopes

Hypothesis

Biotherapeutic sequences that bear analogues to pathogen sequences, may show a strong immunogenicity response in a large proportion of the patient population, as result of a high number of memory T cells that are cross-specific to the biotherapeutic.

➢ New tool needed for the identification of potential cross reactive T cell epitopes.

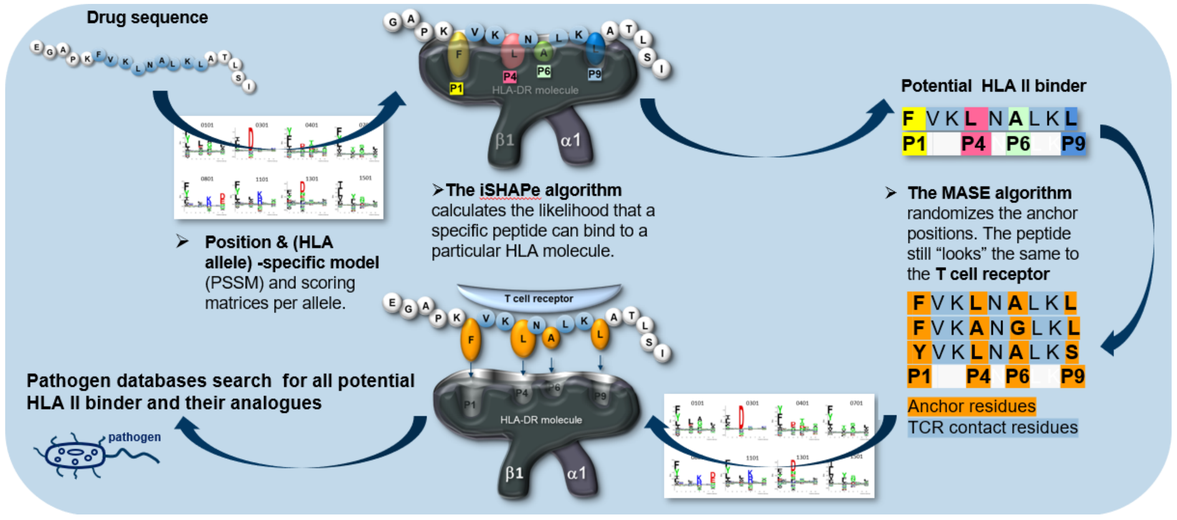


Proof of concept study

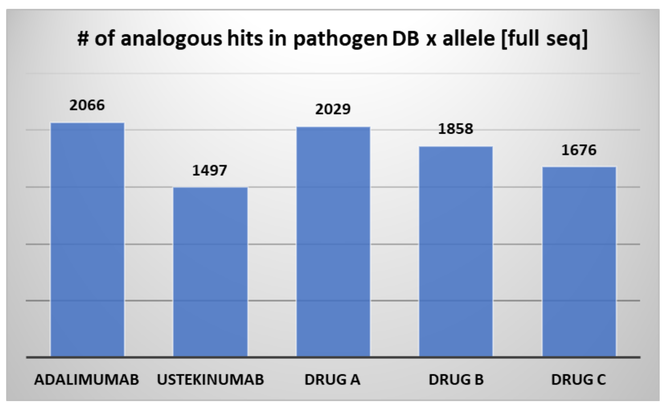
Analysis of drug / pathogen analogy was performed using a newly developed tool called MASE (MAPPs Aggretope Similarity Evaluation). 17 biotherapeutics (1 murine, 2 chimeric, 9 humanized and 5 human antibodies or scFvs) and 21 Databases (19 pathogens, 1 human, 1 CHO) have been included in our preliminary proof of concept study. We will focus on the 5 human antibodies in the following figures. Adalimumab (Humira) and 2 Novartis biotherapeutics (Drug A & B) were selected as examples for antibodies with a relatively high clinical IG rate (% ADA), as well as ustekinumab (Stelara) and Novartis Drug C as examples for a relatively low clinical IG rate.

MASE (MAPPs Aggretope Similarity Evaluation)

MASE is based on the iSHAPe algorithm, exchanging anchor amino acids (P1,4,6,9) of predicted binding peptides with all other 19 natural aa to generate potential analogs that "look alike" to the T cell receptor (anchor amino acids are hidden in the binding pockets of the HLA alleles). Newly generated analogs that are still predicted to bind HLA alleles (2nd iSHAPe run) and original iSHAPe hits are then used in a database look-up to identify potential cross-reactive T cell epitopes.

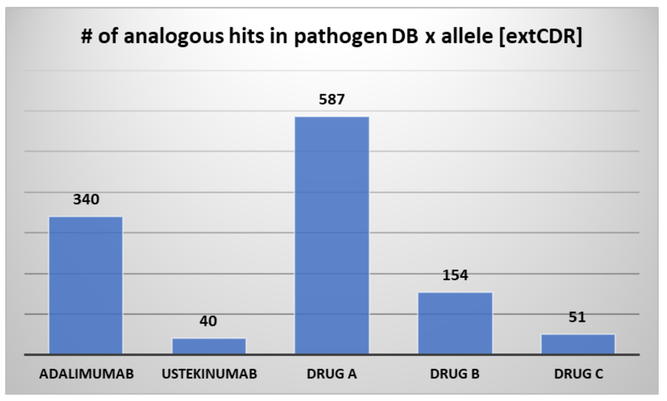


Preliminary Results



Analogous hits in pathogen data

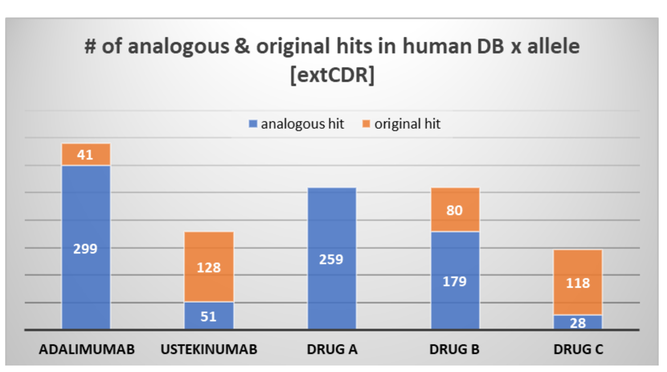
No original hits (original peptide sequence w/o exchanged anchor aa) were found in the 5 tested human antibodies but numerous analogous hits (at least 1 of the 4 anchor aa is exchanged) could be identified. Identical peptides, binding to different alleles were summed up to take the proportion of the population that can present specific hits into account.



CDR overlap ranking

T cell frequencies against foreign sequence regions (CDRs) are expected to be higher than for conserved human sequences

➢ Applying a CDR-based ranking seems to differentiate the drugs better than using the full sequence and correlates better to clinical IG rates (could be wishful thinking – more drugs need to be assessed!)



Humanness ranking

T cell frequencies against conserved human sequence regions are expected to be lower – some kind of tolerance can be expected

➢ Applying only a humanness ranking seems to do little to differentiate the drugs.

drugs	Akkermansia Muciniphila	Bacteroides Fragilis	Bifidobacterium Adolescentis	Bifidobacterium Bifidum	Candidatus Arthromitus	Faecalibacterium Prausnitzii	Salmonella	Neisseria	Clostridium	Campylobacter	Borrelia	Haemophilus influenzae	Corynebacterium diphtheriae	Clostridium tetani	Bordetella pertussis	Streptococcus	Staphylococcus	Shigella	mycobacterium tuberculosis	Vibrio cholerae	Salmonella Typhi	Grand Total
adalimumab	2	3			1	1	4	4	4	6	6	9		3	7	61	1	33	4	11	1	161
ustekinumab					1				1	1		2				5		5	1	1		17
drug A		7	4	3	2	5	8	5	5	12	13	18	11	4	27	63	3	60	23	12	4	289
drug B		5	1		1	1	3	4	7	3	3	7	5	3		26	1	21		6	1	98
drug C			1		1	1		2	1		1	4	6		2	5		6		2	1	33

Hits per pathogen database

Most of the hits are found in Streptococcus and Shigella databases.

➢ Several hits with high similarity to SpeC, a Streptococcus superantigen!

Summary & Outlook

Candidate ranking should be based on a multi-factorial analysis, including not only the # of predicted binders and their humanness but also on hotspot and CDR overlap ranking as well as the analogy of binders to pathogen sequences.

Avoiding sequences that show analogy to pathogens in biotherapeutics development may lower the inherent immunogenicity potential of new drugs dramatically and could give the industry an additional option to ensure patient safety and drug efficacy.

More biotherapeutics will be tested in the POC and *in vitro* assays will be used to substantiate the *in silico* predictions. Analysis of the quality of the hits will be conducted. Especially the hits that can only be found in the high IG candidates vs hits that can only be found in low IG candidates.

Dr. Michael Gutknecht
T +41 795722504; michael.gutknecht@novartis.com

NOVARTIS | Reimagining Medicine