Cross-reactivity of Influenza and ADAMTS13 Peptides in Acquired TTP

Muhamad Rifki Ramadhan

Supervisor: Dr. Can Kesmir
Second reviewer: Prof. dr. Jan Voorberg

Date: 23-02-2022

Abstract

Thrombotic thrombocytopenic purpura (TTP) is a rare but deadly hematologic disease with a prevalence of ~10 cases per million people per year and has a ~90% mortality rate. TTP is caused by low activity of metalloprotease called ADAMTS13, which is responsible for cleaving von Willebrand Factor (vWF), a multimeric glycoprotein that is involved in blood coagulation. Low activity of ADAMTS13 keeps vWF in its multimer state, triggering blood clotting more frequently. In acquired TTP, ADAMTS13 is inhibited by auto-antibodies, but the underlying mechanism of the autoimmunity is still unknown. TTP onsets are frequently reported after infection or vaccination of influenza A virus, especially H1N1 strain. Recent studies suggest that influenza may trigger TTP via molecular mimicry, a process where autoreactive T cells recognize influenza peptides that are cross-reactive with ADAMTS13 peptide. HLA DRB1*11 repeatedly reported as predisposing factor of TTP, while HLA DRB1*04 is considered as a protective factor. Two ADAMTS13 CUB2 domain peptide, FINVAPHAR (position 1324-1332) and LIRDTHSLR (position 1355-1363) were found to bind HLA DRB1*11 and HLA DRB1*03 respectively and can trigger $CD4^+$ T cells response. This research aims to model molecular mimicry between influenza and ADAMTS13 peptide and to predict which influenza peptide can cross-react with ADAMTS13 peptides. We use netMHCIIpan to predict MHC binders and build BLAST-like model to score T cell cross-reactivity between influenza and ADAMTS13 peptide. PS-SCL data on T cell response is incorporated to improve the T cell reactivity prediction. Our model is able to bring up the previously known MHC-binders from ADAMTS13, LIRDTHSLR, as one of the most cross-reactive peptides. It aligns with influenza C peptide VLIADAKGL (nucleoprotein pos. 189-197). We also found several other potentially cross-reactive ADAMTS13 peptides that align with peptides from influenza A and SARS-CoV-2. Furthermore, we applied this model to another environmentally-triggered autoimmune disease, multiple sclerosis. Our model predicts a self-foreign peptide pair that is proven to be cross-reactive in experiments as the most cross-reactive peptide pair. Taken together, our model is able to model molecular mimicry in autoimmune disease and predict the cross-reactive peptides.

Layman’s Summary

Thrombotic thrombocytopenic purpura (TTP) is a rare but deadly blood disease that affect about 10 people per million per year and with mortality rate of about 90%. TTP is caused by the inhibition of a protein called ADAMTS13 by antibodies. Inhibition of ADAMTS13 keeps blood clotting protein vWF in its complex form, triggering blood clot formation more often. Numerous studies reported that TTP emerged after infection or vaccination of influenza A virus. Recent studies suggest that influenza may trigger TTP via molecular mimicry, a phenomenon where a small section of pathogen’s protein, or a pathogen’s peptide, is very similar to the host peptide, so that the immune systems that attack the pathogen’s protein also attack the hosts. In this case, host protein that is attacked is ADAMTS13. HLA DRB1*11, an MHC class II molecule, repeatedly reported as a risk factor for TTP, while HLA DRB1*04 is considered as protective factor. Two ADAMTS13 peptide, FINVAPHAR (position 1324-1332) and LIRDTHSLR (position 1355-1363) were found to bind HLA DRB1*11 and HLA DRB1*03 respectively and can trigger helper T cells response. These findings strengthen molecular mimicry mechanism as the underlying mechanism of TTP. This research aims to model molecular mimicry between influenza and ADAMTS13 peptide and to predict which influenza peptide can cross-react with ADAMTS13 peptides. We use netMHCIIpan to predict MHC binders and build a model to score similarity between influenza and ADAMTS13 peptide. PS-SCL data on T cell response is incorporated to our model to improve the performance. Our model is able to bring up the previously known MHC-binders from ADAMTS13, LIRDTHSLR, as one of the most cross-reactive peptides. It aligns with influenza C peptide VLIADAKGL (nucleoprotein pos. 189-197). We also found several other potentially cross-reactive ADAMTS13 peptides that align with peptides from influenza A and SARS-CoV-2. Furthermore, we applied this model to another environmentally-triggered autoimmune disease, multiple sclerosis. Our model predicts a self-foreign peptide pair that is proven to be cross-reactive in experiments as the most cross-reactive peptide pair. Taken together, our model is able to model molecular mimicry in autoimmune disease and predict the cross-reactive peptides.

Contents

Introduction

Thrombotic Thrombocytopenic Purpura

Thrombotic thrombocytopenic purpura (TTP) is a rare but deadly hematologic disease with a prevalence of ~10 cases per million people per year and has a ~90% mortality rate (Mariotte et al., 2016). Even with an appropriate therapeutic plasma exchange, the mortality rate is still as high as 10-20% (Sadler, 2015). TTP mostly occurs in adults but can also happen in children and adolescents. As its name suggests, TTP is characterized by at least three conditions: thrombosis or blood clotting in small vessels, thrombocytopenia or low count of platelets, and purpura or purple bruises. Other symptoms include fever, weakness, shortness of breath, and headache (Joly et al., 2017).

Pathophysiology of TTP is illustrated in Fig. 1. The formation of thrombi or blood clots in small arterioles is caused by low activity of metalloprotease called ADAMTS13. This protease is responsible for cleaving von Willebrand Factor (vWF), a multimeric glycoprotein that is involved in blood coagulation. In TTP patients, ADAMTS13 has very low activity (less than 10%), causing vWF to stay in its multimer state and triggering blood clotting more frequently. In consequence, blood clots formed in small vessels throughout the body, lowering the number of circulating platelets (Sadler, 2015). Red blood cells passing the clots are subjected to shear stress, leading to rupture of red blood cells within blood vessels, which in turn leads to anemia and schistocyte formation (microangiopathic hemolytic anemia). This condition could also lead to cellular injury and organ damage because of inhibition of blood flow by the clots (Sadler, 2015).

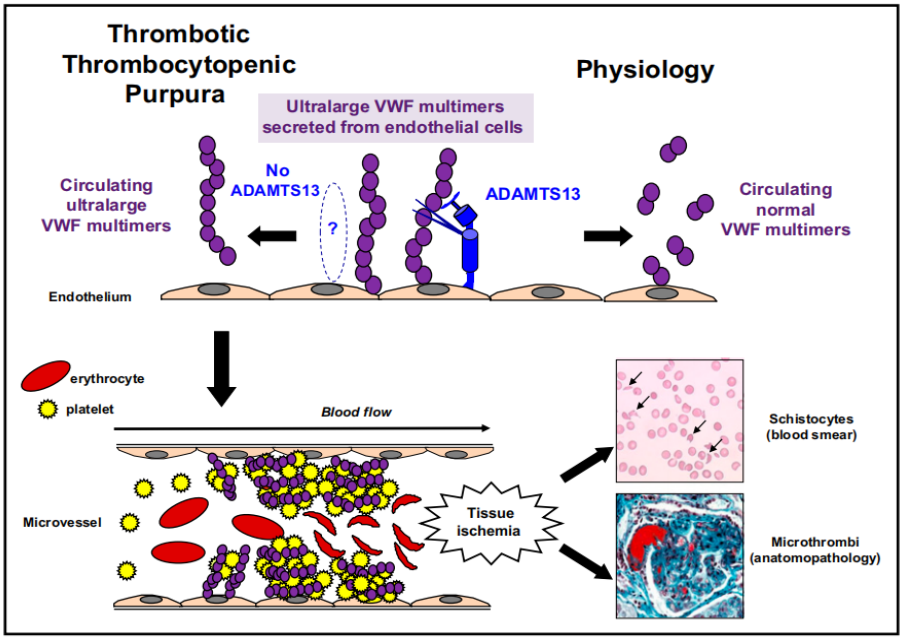


Figure 1. Pathophysiology of TTP (adapted from Joly et al., 2017)

There are two types of TTP based on disease origin: i) Genetic or hereditary TTP is caused by inherited ADAMTS13 deficiency, also known as Upshaw-Schulman syndrome, ii) Acquired TTP, formerly known as idiopathic TTP, is the most common form of TTP which is caused by autoimmunity where antibodies attack ADAMTS13 (Levy et al., 2001). However, the underlying mechanism of the autoimmunity in acquired TTP is still unknown.

Antigen Recognition by T cells

Antigen recognition by T cell receptors is different from antigen recognition by B cell receptors and antibodies. B cell receptors and antibodies bind directly to the surface of intact antigen, while T cell receptors recognize short peptides that are represented by MHC (major histocompatibility complex) molecules. Therefore, epitopes recognized by B cell receptors and antibodies are typically discontinuous amino acid residues that are brought together in the folded protein, while T cell receptor epitopes are short continuous amino acid sequences. These sequences could be anywhere in the protein structure, often buried in the interior of the folded protein. Thus, these sequences can only be recognized by T cell receptors after the protein is unfolded and processed into peptide fragments, which are then presented by MHC molecules (Murphy, 2011).

There are two types of MHC molecules. MHC type I molecules present antigen to $CD8^+$ (cytotoxic) T cells, and MHC type II presents antigen to $CD4^+$ (helper) T cells. Human MHC molecules are also known as HLAs (Human Leukocyte Antigens). MHC class I molecules are expressed by every nucleated cell, while MHC class II molecules are only expressed by antigen-presenting cells (APCs), a group of cells that are involved in cellular immune response, which includes dendritic cells, macrophages, and B cells (Murphy, 2011).

MHC class II molecules consist of two subunits, α and β, that are not covalently bound. Each chain consists of two domains. The α_2 and β_2 domains are membrane-bound while α_1 and β_1 domains form the peptide-binding cleft. This structure is similar to MHC class I molecules, which consist of three α chain domains and one β_2m chain domain, and only the αα chain is bound to the membrane and form the peptide-binding cleft. Except for β_2m chain of MHC class I, the MHC subunit genes are highly polymorphic, and the major differences between different alleles are in the peptide-binding cleft, influencing peptide-binding specificity (Murphy, 2011). The MHC class II molecule and peptide-binding cleft are depicted in Fig. 2.

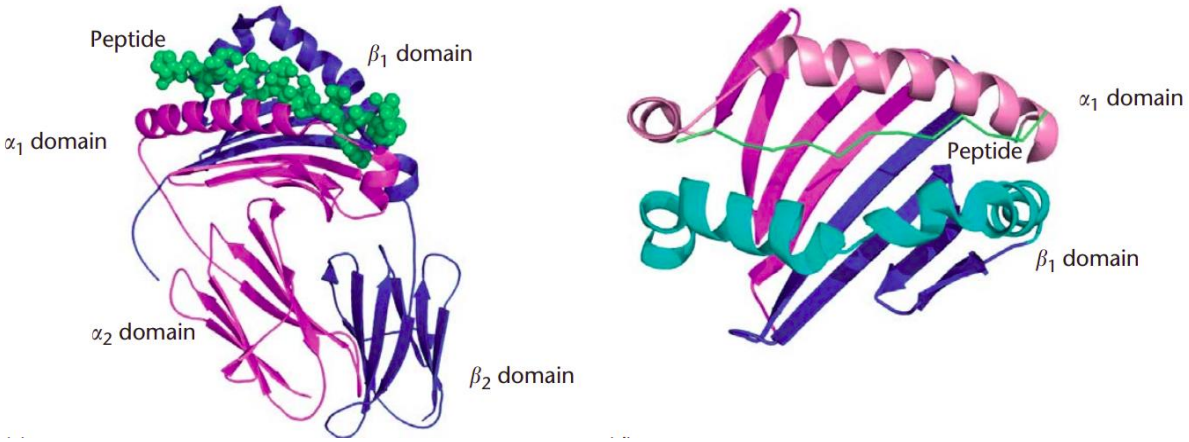


Figure 2. The structure of MHC class II molecule (left) and MHC class II binding cleft (right) (adapted from Liu and Gao, 2011)

Because of the different peptide-binding cleft structure, peptide-binding of MHC class II molecules differs in several ways compared to MHC class I. Peptides bound to MHC class I molecules are short peptides of 8-10 amino acids long, whereas peptides bound to MHC class II are at least 13 amino acids long and can be much longer. This is because the binding groove of MHC class II molecules is not closed in both ends, allowing peptides to protrude beyond the binding groove, as opposed to being completely buried in MHC class I molecules (Murphy, 2011).

Part of the peptide that lies within the groove of MHC class II molecules is called the core peptide and is 9 amino acids long. Structural studies show that the side chains of amino acid residues 1, 4, 6, and 9 are facing towards and bound to MHC binding groove residues. These residues in turn determine MHC binding specificity and are called anchor residues (Reynisson et al., 2020). T cell receptor interacts with the peptide-MHC complex, making contact both with the peptide and the MHC molecules (Harkiolaki et al., 2009).

Molecular Mimicry and Cross-Reactivity

Molecular mimicry was first described by Damian in 1964 as the existence of foreign antigens that are similar to host antigens that may facilitate the infectious agents to avoid host immune responses (Damian, 1964). Almost at the same time, the evidence of immunological cross-reactivity has been observed by Kaplan et al. (1962) between group-A streptococcal cells and human heart tissue.

In 1983, Fujinami et al. found that murine antibodies against measles virus and herpes simplex virus (HSV) could be reactive to human cells. Two years later, they demonstrated that rabbits sensitized by hepatitis B virus polymerase (HBVP) developed encephalomyelitis, an autoimmune disease that attack Myelin basic protein (MBP) in central nervous system. They found that rabbits injected with HBVP peptide produce antibody that reacts with HBVP itself and also with MBP (Fujinami & Oldstone, 1985). These findings reflect the role of molecular mimicry and cross-

reactivity between foreign and self antigens in the onset of autoimmune diseases. Over the years, more autoimmune diseases were discovered to be triggered by molecular mimicry and cross-reactivity, including multiple sclerosis (MS), antiphospholipid syndrome (APS), and Sjogren's syndrome (SS) that have been associated with the infection of *M. avium*, *R. intestinalis*, and Epstein-Barr virus (EBV) respectively (Rojas et al., 2018). Acquired TTP is considered to be initiated by molecular mimicry and cross-reactivity and belongs in this autoimmune disease group.

Molecular mimicry involves proliferation of autoreactive T cells (Fig. 3). These autoreactive T cells escaped negative selection in the thymus because of their low affinity to host antigen or absence of presentation of self antigen in the thymus. These T cells will not cause autoimmunity by itself because of their low affinity to self antigens and their low abundance. However, once they are stimulated by a foreign antigen, they will proliferate to a high number, activating B cells, and can damage the host cells expressing the self antigen. Over time, such B cells undergo affinity maturation that increases their affinity to the host antigen, worsening the disease progression (Rojas et al., 2018).

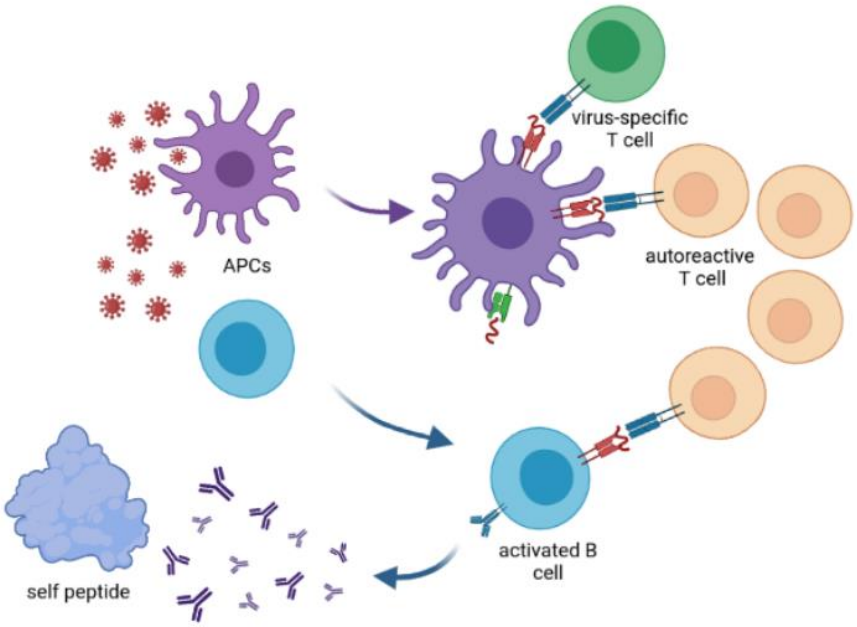


Figure 3. Illustration of molecular mimicry (self illustration, created with biorender.com)

Initially, T cell response was assumed to be very specific, which makes it less likely for cross-reactivity to occur. However, recent studies suggest that while there are critical residues of a peptide that determine their binding specificity to MHC molecule and T cell receptor, there is a degree of plasticity in the other residues. This allows the possibility of the same T cell clone to respond against multiple antigens, including self antigens. However, the unknown nature of peptide-MHC binding and antigen recognition by T cell receptors, especially MHC II and $CD4^+$ T cell binding, makes it difficult to predict cross-reactivity between antigens.

Risk Factors of Acquired TTP

The association between TTP and influenza infection has been well known for a number of years. Numerous studies reported that TTP emerged after infection or vaccination of influenza A virus, especially H1N1 strain (Mamori et al., 2008; Dias et al., 2009; Hermann et al., 2010; Lee et al., 2011; Koh et al., 2012; Jonsson et al., 2015).

Kosugi et al. (2010) found that influenza A infection triggers the production of anti-ADAMTS13 IgGs. This is an important finding because B cells that are triggered by an antigen will only produce antibody isotype IgM, and only switch to IgG after being activated by $CD4^+$ T cells which also detect the same antigen. Therefore, the existence of anti-ADAMTS13 IgGs indicates that $CD4^+$ T cells are involved in the onset of acquired TTP. ADAMTS13 auto-antibodies are also increased in titer over time and gain inhibitory capacity (Froehlich-Zahnd, 2012). This implies affinity maturation, a process also dependent on $CD4^+$ T cells. Taken together these data suggest that influenza may trigger TTP via molecular mimicry where autoreactive $CD4^+$ T cells recognize influenza peptides that are cross-reactive with ADAMTS13 peptide.

HLA DRB1*11 allele is a very well-studied predisposing factor of TTP. Several studies reported HLA DRB1*11 over-representation in TTP patients compared to control population. On the other hand, HLA DRB1*04 is considered a protective factor for TTP based on its under-representation in TTP patients (Scully et al., 2010; Coppo et al., 2010; John et al., 2011). One study found that HLA DRB1*14 and DRB1*15 are also predisposing factors of TTP while HLA DRB1*07 and DRB1*13 are protective (Sinkovits et al., 2017). Study on Japanese population revealed a very different TTP-HLA association, with DRB1*0803, DRB3/4/5*blank, DQA1*0103, and DQB1*0601 considered as susceptible alleles and DRB1*1501 and DRB5*0101 as protective ones (Sakai et al., 2020).

Sorvillo et al. (2013) demonstrated that ADAMTS13 CUB2 domain peptide FINVAPHAR (position 1324-1332) can be presented by dendritic cells with HLA DRB1*1101 molecule in vitro. They also found that LIRDTHSLR (position 1355-1363), another peptide from CUB2 domain, is presented by HLA DRB1*0301 molecule. Verbij et al. (2016) identified FINVAPHAR-reactive $CD4^+$ T cells in HLA DRB1*11-positive patients, while ASYILIRD-reactive (peptide four residues upstream of LIRDTHSLR) $CD4^+$ T cells were identified in HLA-DRB1*03-positive patients. These findings confirm the autoreactive $CD4^+$ T cells involvement and corroborate the molecular mimicry mechanism in the onset of acquired TTP.

Nevertheless, we still do not know much about molecular mimicry in acquired TTP, including which influenza peptide can trigger the cross-reactivity. This research aims to model molecular mimicry mechanism between influenza and ADAMTS13 peptide and to predict which influenza peptide can cross-react with which ADAMTS13 peptide.

Data and Methods

Obtaining sequences

ADAMTS13 sequence was obtained from Uniprot (entry: Q76LX8). Isoform 1 was used because it is canonical and has the longest sequence. Influenza virus sequences were also obtained from Uniprot, which includes influenza A virus H9N2 (Proteome ID: UP000105043), H3N2 (Proteome ID: UP000115734), H5N1 (Proteome ID: UP000131152), and H1N1 (Proteome ID: UP000009255), influenza B virus (Proteome ID: UP000008158), and influenza C virus (Proteome ID: UP000008286). Human immunodeficiency virus 1 (HIV-1) (Proteome ID: UP000002241), hepatitis C virus (HCV) (Proteome ID: UP000000518), and human parvovirus B19 (HPV B19) (Proteome ID: UP000006624) are used as control. Novel coronavirus SARS-CoV-2 (Proteome ID: UP000464024) is also included in this analysis.

Peptide-MHC binding predictions

To predict peptide-MHC binding, we used NetMHCIIpan-4.0, a tool that can predict peptide binding to any MHC II molecule of known sequence using Artificial Neural Networks (ANNs). This tool uses eluted ligand data, which implicitly contains information from steps of MHC II antigen presentation such as antigen digestion, ligands loading, and transport (Reynisson et al., 2020). We performed the peptide binding predictions for HLA DRB1*1101 and DRB1*0301 allele, the two HLA alleles which are known to present ADAMTS13 peptides (FINVAPHAR and LIRDTHSLR, respectively) and trigger $CD4^+$ T cell response. The peptide length is set to 15, as the length of the known peptide (Sorvillo et al., 2013; Verbij et al., 2016).

Top 10% peptides with the best eluted ligand score are selected. Peptides that are overlapping with other peptides with higher affinity are eliminated. The remaining peptides are then called the top binders.

Cross-reactivity scoring model

The basic concept of the similarity/cross-reactivity scoring we used is similar to that of BLAST. Each self HLA-binding peptide is aligned against every foreign peptide that is predicted to bind the same HLA molecule, and a cross-reactivity score is calculated based on their sequence similarity. PMBEC matrix, a substitution matrix optimized for protein-peptide interaction (Kim et al., 2009), is used to score the similarity between amino acids of the aligned peptides. The cross-reactivity score of each self-foreign peptide alignment, *CR*, is a sum of PMBEC scores from each amino acid aligned, s_i, normalized with respect to the maximum score, s_{max_i}. The maximum score is the score of an amino acid when aligned with itself.

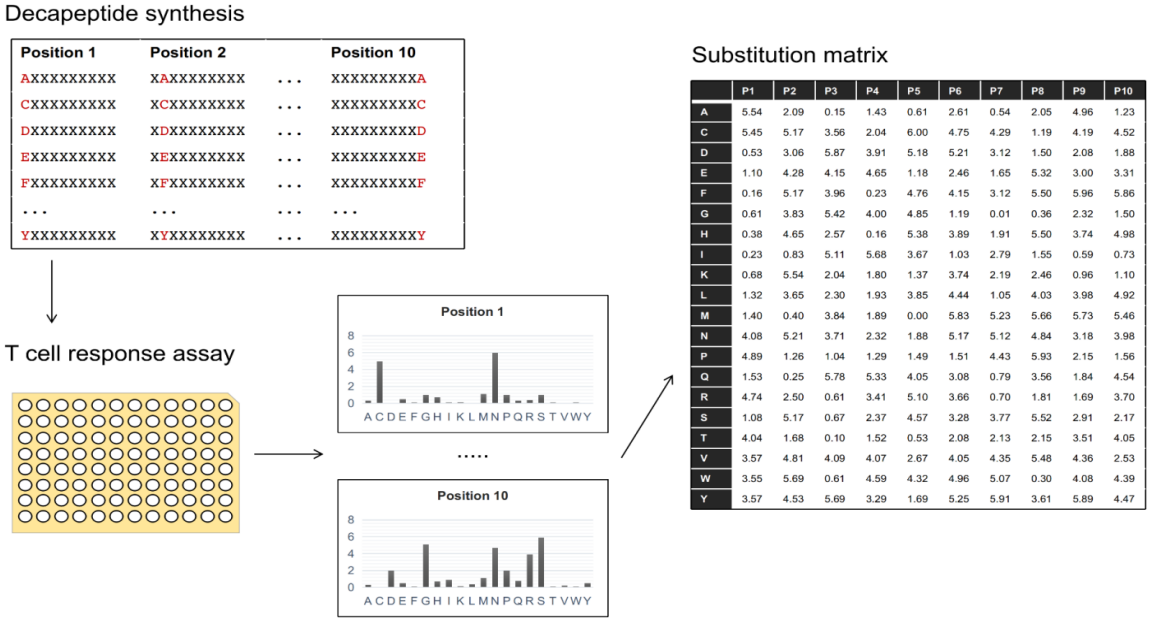
We further optimize our model by differentiating our scoring using the peptide-TCR contact positions previously described by Wang & Reinherz (2002): Peptide positions P2, P3, P5 and P8 are considered as TCR contact residues. We make three variations of cross-reactivity models: Model 1 only scores the defined TCR contact residues ($w_2 = w_3 = w_5 = w_8 = 1$; $w_1 = w_4 = w_6 = w_7 = 0$), Model 2 also scores other residues in core peptide but with 0.5 weight ($w_1 = w_4 = w_6 = w_7 = 0.5$), and Model 3 puts 0.75 weight ($w_1 = w_4 = w_6 = w_7 = 0.75$) on other core peptide residues. Thus, the final model becomes:

$$CR\ score = \sum_{i=1}^{9} w_i \frac{s_i}{s_{max_i}} \qquad \textit{Equation 1}$$

This cross-reactivity scoring model, as well as the following model, is available through github (https://github.com/mrramadhan21/ttp_crossreactivity).

Diversity-based cross-reactivity scoring model

Several studies have tried to determine TCR specificity using positional scanning substrate combinatorial library (PS-SCL). PS-SCL is a synthetic mixture of peptides with a specific amino acid fixed in one or more positions, with the remaining positions being randomized with a mixture of equimolar amino acids (Fig. 4). To determine T cell specificity, a T cell response assay is performed using the mixture of the libraries, APCs, and the T cell clones. The benefit of this method is the effect of the fixed amino acid in a particular position can be determined independent of the other residues. T cell activity data is then normalized and used to construct a position-based amino acid substitution matrix (Fig. 4)



	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
A	5.54	2.09	0.15	1.43	0.61	2.61	0.54	2.05	4.96	1.23
C	5.45	5.17	3.56	2.04	6.00	4.75	4.29	1.19	4.19	4.52
D	0.53	3.06	5.87	3.91	5.18	5.21	3.12	1.50	2.08	1.88
E	1.10	4.28	4.15	4.65	1.18	2.46	1.65	5.32	3.00	3.31
F	0.16	5.17	3.96	0.23	4.76	4.15	3.12	5.50	5.96	5.86
G	0.61	3.83	5.42	4.00	4.85	1.19	0.01	0.36	2.32	1.50
H	0.38	4.65	2.57	0.16	5.38	3.89	1.91	5.50	3.74	4.98
I	0.23	0.83	5.11	5.68	3.67	1.03	2.79	1.55	0.59	0.73
K	0.68	5.54	2.04	1.80	1.37	3.74	2.19	2.46	0.96	1.10
L	1.32	3.65	2.30	1.93	3.85	4.44	1.05	4.03	3.98	4.92
M	1.40	0.40	3.84	1.89	0.00	5.83	5.23	5.66	5.73	5.46
N	4.08	5.21	3.71	2.32	1.88	5.17	5.12	4.84	3.18	3.98
P	4.89	1.26	1.04	1.29	1.49	1.51	4.43	5.93	2.15	1.56
Q	1.53	0.25	5.78	5.33	4.05	3.08	0.79	3.56	1.84	4.54
R	4.74	2.50	0.61	3.41	5.10	3.66	0.70	1.81	1.69	3.70
S	1.08	5.17	0.67	2.37	4.57	3.28	3.77	5.52	2.91	2.17
T	4.04	1.68	0.10	1.52	0.53	2.08	2.13	2.15	3.51	4.05
V	3.57	4.81	4.09	4.07	2.67	4.05	4.35	5.48	4.36	2.53
W	3.55	5.69	0.61	4.59	4.32	4.96	5.07	0.30	4.08	4.39
Y	3.57	4.53	5.69	3.29	1.69	5.25	5.91	3.61	5.89	4.47

Figure 4. Schematic diagram of PS-SCL T cell response experiment

From the substitution matrices of PS-SCL studies, we calculated the amino acid's Simpson's diversity index on each position. Simpson's diversity index is a value that is usually used in ecology to measure species diversity (Begon et al., 2006). Simpson's diversity index can be derived from abundance of a particular species, n_i, and the total abundance of all species, *N*, as follows:

$$D = 1 / \sum_{i=1}^{n} \left(\frac{n_i}{N}\right)^2 \qquad \textit{Equation 2}$$

Our implementation of the diversity index ranges from 1 to 20 and indicates how many amino acids can trigger T cell response in a particular position. If a position has a low diversity index, that means only a few amino acids in that position can produce T cell response, thus the position is most likely to be important in pMHC-TCR interaction. Likewise, if a position has a high diversity index, it allows a lot of amino acid substitution, and is most likely not important in pMHC-TCR interaction.

We implement this diversity index as a number of allowed amino acid substitutions for each position in our cross-reactivity scoring model. For instance, a position that has a diversity index of 2, can only be substituted by the two most favorable amino acids based on the substitution matrix score. For all other substitutions, a negative value of -3, which is an arbitrary number lower than the lowest score in the matrix, is added to the alignment score as a penalty. Positions that are substituted by the "allowed" amino acids will get their respective substitution matrix scores. In this model, substitution matrix used is BLOSUM100 to impose starker difference between favorable and unfavorable amino acids.

Weights are added to each position to avoid peptide alignment results that have very high scores just because they have exact amino acid alignments in less important positions. We define this weight in such a way that more diverse, i.e., flexible positions get automatically less weight in the calculation of the cross-reactivity score:

$$w_i = \frac{1}{\sqrt{D_i}} \qquad \textit{Equation 3}$$

Multiple sclerosis (MS) analysis

The myelin basic protein (MBP) sequence is obtained from Uniprot (entry: P02686), as are *Escherichia coli* (Proteome ID: UP000000558) and *Mycobacterium avium* proteome (Proteome ID: UP000019908). The peptide-MHC binding prediction and cross-reactivity scoring prediction are done in the same way as TTP analysis (peptide length 15, threshold 10%).

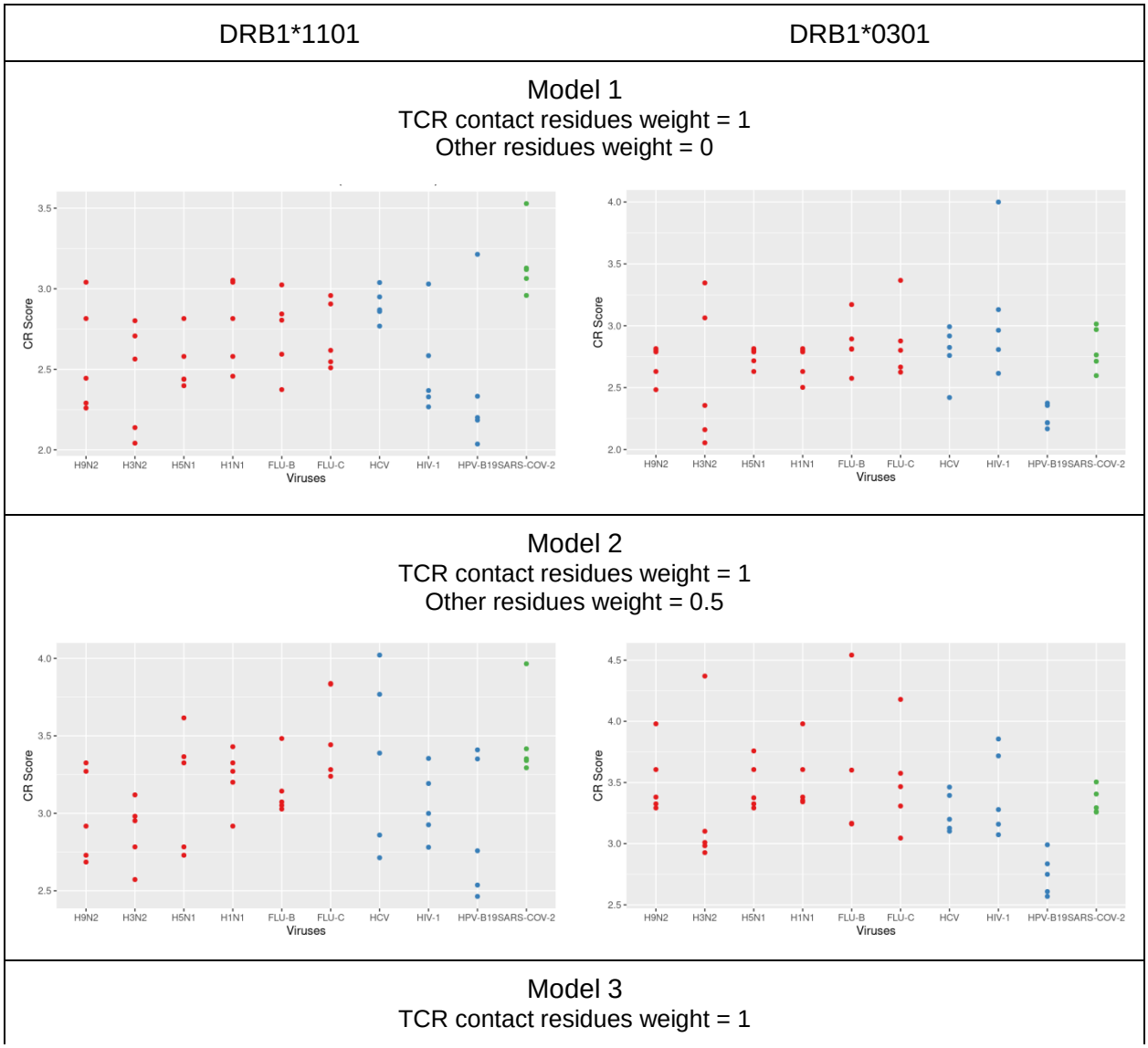
Because of the size of *E. coli* and *M. avium* proteome, we did not include the whole proteome in the analysis. Instead, an antigen ensemble is constructed based on the presence of the epitopes

originated from these bacteria in the IEDB database. We looked at the registered B and T cells epitopes of *E. coli* and *M. avium*, and selected proteins from the proteome data in which the epitopes belonged to build the antigen ensemble.

Results

Cross-reactivity scoring model

Figure 5 shows the prediction results of the three models. In Model 1, there is no ADAMTS13-influenza peptide alignment that indicates strong cross-reactivity (Fig. 5 and Table 1a). Most of the influenza peptides fall on relatively similar score both in HLA DRB1*1101 and DRB1*0301 alleles. The highest score actually comes from SARS-CoV-2 peptide in HLA DRB1*1101 and HIV-1 in DRB1*0301 (Table 1a). The experimentally identified peptide FINVAPHAR did not have any high scoring alignment with any influenza peptides in HLA DRB1*1101, and LIRDTHSLR is not predicted to have a significant cross-reactivity with any influenza peptides in HLA DRB1*0301.



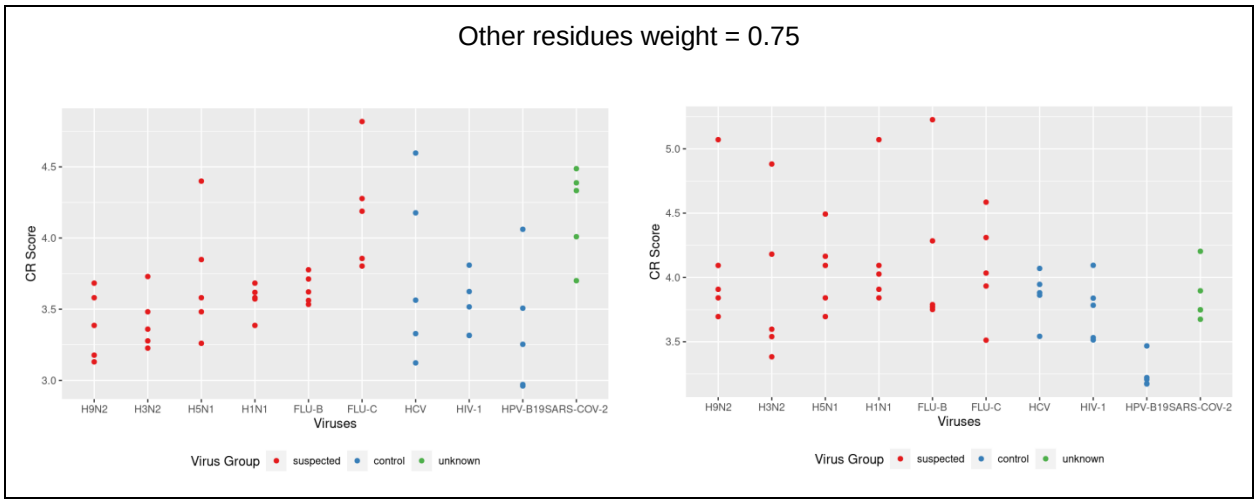


Figure 5. Cross-reactivity score of ADAMTS13 peptides with various influenza viruses and control viruses. Shown in the graph are the five best cross-reactivity scores of each virus.

Table 1a. Top 5 cross-reactivity score of ADAMTS13 peptide using Model 1

Self peptide core	Virus peptide core	Score	Pos in ADAMTS13	Virus	Protein	Position
DRB1*1101						
YVLTNLNIG	VILLNKHID	3.528	100	SARS-COV-2	P0DTC9	347
WKVMSLGPC	YKVFSPAAS	3.213	1013	HPV-B19	Q6TV11	278
YVLTNLNIG	YVLMDGSII	3.128	100	SARS-COV-2	P0DTD1	2951
LKAQASLRG	LKSIAATRG	3.119	1392	SARS-COV-2	P0DTD1	4965
WKVMSLGPC	FKVNSTLEQ	3.063	1013	SARS-COV-2	P0DTD1	5667
DRB1*0301						
FLKAQASLR	YLKDQQLLG	4.000	1391	HIV-1	P04578	582
YIANHRPLF	IIANKRMLE	3.366	581	FLU-C	Q6I7C4	57
ITEDTGFDL	YVEDTKIDL	3.346	207	H3N2	B2ZV32	425
LLEDGRVEY	ILPTGRVEH	3.171	617	FLU-B	P03474	260
GSVDEKLPA	ISPIETVPV	3.130	843	HIV-1	P04585	586

Using the second model, there is still no obvious highly cross-reactive peptide in DRB1*1101 (Fig. 5). For DRB1*0301, there are several ADAMTS13-influenza peptide alignments that score high compared to control viruses. The highest score is from influenza B peptide ILPTGRVEH

(neuraminidase position 260-268) that is aligned with ADAMTS13 peptide LLEDGRVEY (pos. 617-625). The second and third highest score are from influenza A H3N2 (YVEDTKIDL, HA protein pos. 425-433) and influenza C (IIANKRMLE, PB2 protein pos. 57-65) that are aligned with ADAMTS13 peptide ITEDTGFDL (pos. 207-215) and YIANHRPLF (pos. 581-589) respectively. The influenza A H3N2 peptide YVEDTKIDL is reported as an immunodominant epitope for T cell recognition (Gelder et al., 1995). LIRDTHSLR, the peptide that is known to bind HLA DRB1*0301 and can trigger CD4+ T cell reaction, is aligned with peptide LRRDQKSLR from influenza A and has the fourth best score, but its score is similar to that of control viruses (Table 1b).

Table 1b. Top 5 cross-reactivity score of ADAMTS13 peptide using Model 2

Self peptide core	Virus peptide core	Score	Pos in ADAMTS13	Virus	Protein	Position
DRB1*1101						
LVPHEEAAA	YVPESDAAA	4.021	1135	HCV	P27958	1931
LKAQASLRG	LKSIAATRG	3.965	1392	SARS-COV-2	P0DTD1	4965
LKAQASLRG	FKRTASQRA	3.837	1392	FLU-C	Q6I7C4	746
FTYFQPKPR	FTYFAPKQG	3.833	672	FLU-C	Q6I7C4	639
WKVMSLGPC	YKVLVLNPS	3.768	1013	HCV	P27958	1247
DRB1*0301						
LLEDGRVEY	ILPTGRVEH	4.541	617	FLU-B	P03474	260
ITEDTGFDL	YVEDTKIDL	4.370	207	H3N2	B2ZV32	425
YIANHRPLF	IIANKRMLE	4.179	581	FLU-C	Q6I7C4	57
LIRDTHSLR	LRRDQKSLR	3.979	1355	H9N2	Q9DHF7	32
LIRDTHSLR	LRRDQKSLR	3.979	1355	H1N1	P03496	32

In Model 3, the influenza cross-reactivity scores become higher and more distinct than other viruses in DRB1*0301. The LIRDTHSLR / LRRDQKSLR alignment is now the second-best alignment with a clear difference with other viruses. The other three alignments mentioned before in Model 2 still have distinctly high scores. In contrast, we still do not see FINVAPHAR peptide within the most cross-reactive peptides for DRB1*1101, and even though the highest cross-reactivity score comes from influenza, the score difference with other viruses is not convincing (Table 1c).

Given these results, model 3 seems to be the best model so far, at least for DRB1*0301: it can distinguish the difference between influenza and other viruses, while it also scores the known cross-reactive peptide, LIRDTHSLR, amongst the highest. However, we still need to fine tune this model further to get a better result for DRB1*1101.

Table 1c. Top 5 cross-reactivity score of ADAMTS13 peptide using Model 3

Self peptide core	Virus peptide core	Score	Pos in ADAMTS13	Virus	Protein	Position
DRB1*1101						
FTYFQPKPR	FTYFAPKQG	4.818	672	FLU-C	Q6I7C4	639
LVPHEEAAA	YVPESDAAA	4.597	1135	HCV	P27958	1931
LTPIAAVHG	LKSIAATRG	4.487	374	SARS-COV-2	P0DTD1	4965
FTYFQPKPR	ITHFQRKRR	4.400	672	H5N1	Q9Q0V0	179
LKAQASLRG	LKSIAATRG	4.388	1392	SARS-COV-2	P0DTD1	4965
DRB1*0301						
LLEDGRVEY	ILPTGRVEH	5.226	617	FLU-B	P03474	260
LIRDTHSLR	LRRDQKSLR	5.071	1355	H9N2	Q9DHF7	32
LIRDTHSLR	LRRDQKSLR	5.071	1355	H1N1	P03496	32
ITEDTGFDL	YVEDTKIDL	4.882	207	H3N2	B2ZV32	425
YIANHRPLF	IIANKRMLE	4.585	581	FLU-C	Q6I7C4	57

Improving the cross-reactivity model: TCR-pMHC interaction prediction using PS-SCL data

The previous results suggest that there is room for improving the predictions of TCR and peptide-MHC interaction. Clearly, some positions of the peptide are more important than the others (Harkiolaki et al., 2009), but there is not a clear quantitative model to implement this. In order to build such a model, we collected data from the combinatorial peptide library studies on TCR response. These studies analyze TCR-pMHC interaction by synthesizing peptides with different amino acids on a particular position, and leaving the other positions randomly synthesized. This library of peptides is called positional scanning substrate combinatorial library (PS-SCL) (Schneider et al., 2009).

We obtained substitution matrix data from two studies (Birnbaum et al., 2014; Zhao et al., 2001). Birnbaum et al. used Ob.1A12 and Ob.2F3 T cell clones which are reactive to peptides displayed by HLA DRB1*1501 molecules, while Zhao et al. used HLA DRB1*0404-reactive T cell clone GP5F11. The substitution matrices from these studies are displayed in Supplementary Table 1a and 1b.

Simpson's diversity index (Begon et al., 2006) was calculated for each position from the substitution matrix data. This diversity measure is usually used in ecology to measure species diversity. Our Simpson's diversity index ranges from 1 to 20, representing the average number of amino acids present in that position. More details on the exact implementation of this model are given in Data and Methods.

Simpson's diversity index across the peptide is shown in Fig. 6. Both data agree that P2, P3, and P5 play an important role in TCR contact because they have relatively low amino acid diversity, indicating that only few mutations are allowed for TCR binding, as previously expected. However, it seems like our previous model underestimates the importance of P1 and P4, as well as P6 and P7. P1 and P4 turn out to be as important as P2, P3, and P5, especially in DRB1*0404 restricted T cell clones, and while P6 and P7 do not have a diversity index as low as P1-P5, they are not totally generic either. Interestingly, DRB1*1501 and DRB1*0404 restricted T cell clones show opposite trends in permeability at positions 8 and 9. Contrary to the literature (Harikolaki et al., 2009) and DRB1*0404 restricted T cell clone data, DRB1*1501 restricted T cell clone data indicates that P9 is more important than P8.

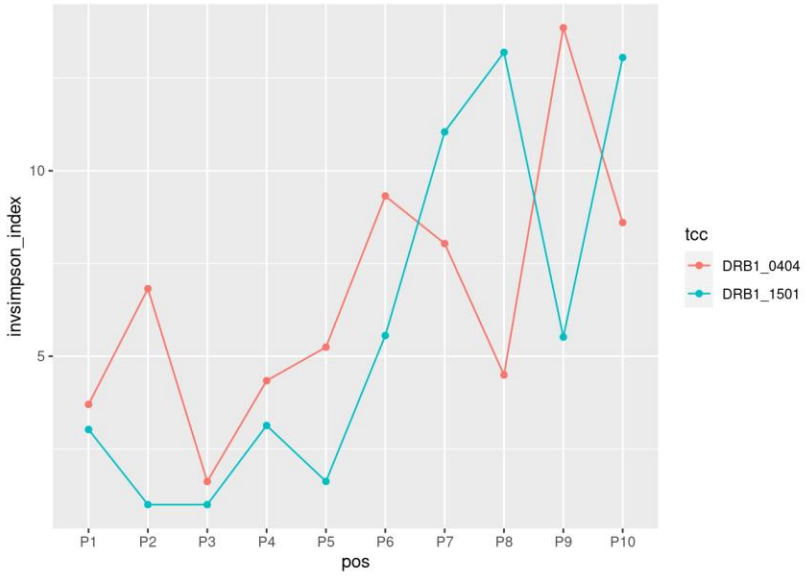


Figure 6. Simpson diversity index of the DRB1*0404 (pink) and DRB1*1501 (blue) restricted T cell clones

In summary, PS-SCL data suggests that the amino acid specificity in the peptide-TCR interaction is much less flexible than it was initially thought. Several positions allow only very few amino acid substitutions. This means that the simple matrix scoring that we used before is too tolerant for

substitutions between distantly related amino acids. Therefore, we improve the cross-reactivity scoring model (Eq. 1) by implementing the diversity measure plotted in Fig. 6 as the number of allowed amino acid substitutions for each position. For example, position 5 in DRB1*1501 restricted T cells has a diversity index of 2, which we interpret as this position can only be substituted by the two most favorable amino acids based on the BLOSUM matrix. For all other substitutions, a negative value of -3 is added to the binding score as a penalty. This value is picked to be just lower than the lowest substitution score in the matrix. Positions that are substituted by the "allowed" amino acids will get their respective BLOSUM score.

We then add weights to each position to avoid peptide alignments having a very high scores just because they have exact amino acid alignment in less important positions. This weight is also based on the diversity index (Fig. 6, Eq. 3). The final scoring then becomes:

$$CR\ score\ = \sum_{i=1}^{9} \frac{1}{\sqrt{D_i}} \cdot \frac{s_i}{s_{max_i}} \qquad \textit{Equation 4}$$

Diversity-based cross-reactivity scoring model

From the two available alleles on which we have the PS-SCL data, DRB1*0404 is found to have protective association with TTP (Scully et al., 2010; Coppo et al., 2010; John et al., 2011), while DRB1*1501 is found to be a TTP predisposing factor in one study (Sinkovits et al., 2017), but not the other (Scully et al., 2010; Coppo et al., 2010; John et al., 2012). Unfortunately, substitution matrix data for DRB1*1101 restricted T cell clone, the allele with strong susceptibility to TTP, is not available.

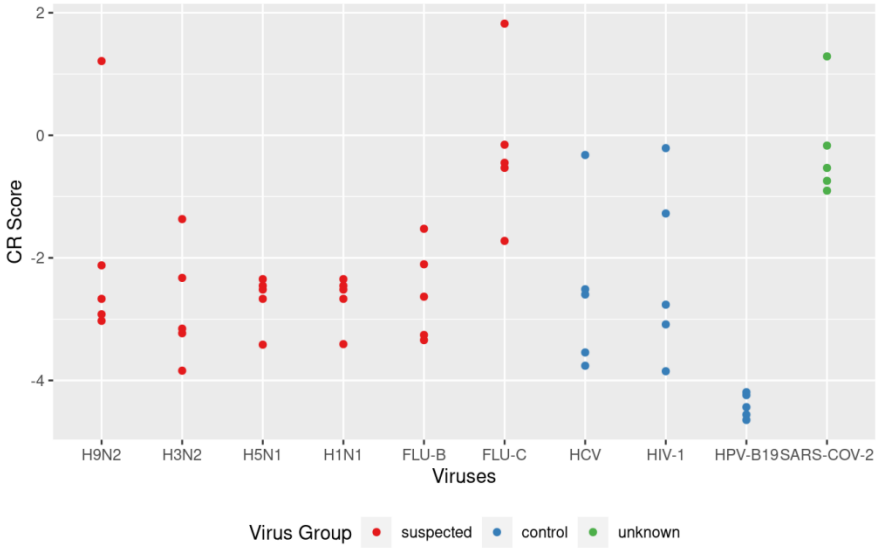


Figure 7. Cross-reactivity score of ADAMTS13 peptide with diversity model (DRB1*1501)

Figure 7 shows the result of cross-reactivity score analysis using DRB1*1501 data and the new diversity-based cross-reactivity model (Eq. 4). We can clearly see that there are three peptide alignments with especially high cross-reactivity scores: two from influenza and one from SARS-CoV-2 (Table 2 and Fig. 7). The highest score comes from influenza C peptide VLIADAKGL (nucleoprotein pos. 189-197), which is aligned with ADAMTS13 peptide ILIRDTHSL (pos. 1355-1363). As mentioned earlier, this ADAMTS13 peptide has been proved to bind to HLA DRB1*0301 molecule (Sorvillo et al., 2013) and can trigger $CD4^+$ T cell response (Verbij et al., 2016). This result suggests that this peptide may also bind with HLA DRB1*1501 molecule, and also suggests a foreign peptide candidate to which it can cross-react: VLIADAKGL from influenza C.

Table 2. Top 5 cross-reactivity score of ADAMTS13 peptide using diversity-based prediction (DRB1*1501)

Self peptide core	Virus peptide core	Score	Pos in ADAMTS13	Virus	Protein	Position
ILIRDTHSL	VLIADAKGL	1.823	1356	FLU-C	Q6I7C0	188
LLRYGSQLA	ILRLGSPLS	1.288	1278	SARS-COV-2	P0DTD2	41
LVLYITRFD	PVLYINVAD	1.212	171	H9N2	Q9ICY2	296
VSSYLSPGA	IFSFLATAA	-0.153	43	FLU-C	Q6I7C2	149
LLRYGSQLA	ILRKGGRTI	-0.167	1278	SARS-COV-2	P0DTD1	393

The second influenza peptide alignment with a distinctly high score is PVLYINVAD (influenza A H9N2 neuraminidase pos. 296-304) that is aligned with ADAMTS13 peptide LVLYITRFD (pos. 171-179) (Table 2). This is a novel ADAMTS13 peptide which has not been reported in HLA binding studies before. The fact that this is an influenza A peptide is a good indication for an association to TTP, since most of the influenza-TTP association studies reported influenza A as the trigger of TTP and Kosugi et al. (2010) found that influenza A infection can trigger anti-ADAMTS13 IgG production.

Also, within SARS-CoV-2 proteome, we found a peptide with specially high cross-reactivity score: ILRLGSPLS (ORF9b protein pos. 41-49) that is aligned with LLRYGSQLA (pos. 1279-1287) of ADAMTS13 (Table 2). This is a very interesting result since there are some reported onsets of TTP after SARS-CoV-2 infection (Capecchi et al., 2020; Cohen et al., 2021). Our predictions support these reports that SARS-CoV-2 infection may trigger TTP via molecular mimicry.

Nevertheless, the known ADAMTS13 DRB1*1101-restricted peptide FINVAPHAR still has not emerged as a high-scoring peptide. The slightly upstream peptide LFINVAPHA is among the peptides that pass the MHC binding threshold, but there is no influenza peptide that is similar enough with it to have a high cross-reactivity score. This may not be the case if we use DRB1*1101 PS-SCL data, but unfortunately the data is not available.

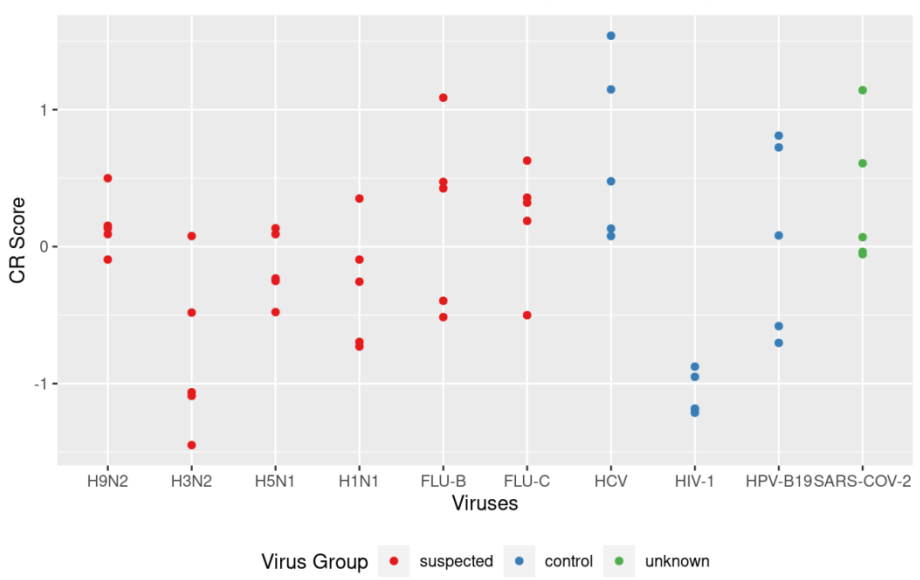


Figure 8. Cross-reactivity score of ADAMTS13 peptide with diversity model (DRB1*0404)

Figure 8 shows the results of TTP cross-reactivity analysis using DRB1*0404 data. Unlike in the DRB1*1101 result, there is hardly any influenza peptide alignments that score considerably higher than the others (Fig. 8 and Table 3). The two highest scores come from HCV, our control virus. SARS-CoV-2 has again a high scoring peptide alignment: IVYTACSHA (replicase polyprotein 1ab pos. 5625-5633) aligned with ILYCARAHG (ADAMTS13 pos. 971-979). The highest influenza score comes from influenza B: LSGMGTTAT (matrix protein 1 pos. 80-88) aligned with LCGLASKPG (ADAMTS13 pos. 929-937). These results suggest that it is unlikely for ADAMTS13-influenza cross-reactivity to happen in HLA DRB1*0404, which is in accordance to the studies that reported HLA DRB1*04 as a protective factor against TTP.

Table 3. Top 5 cross-reactivity score of ADAMTS13 peptide using diversity-based prediction (DRB1*0404)

Self peptide core	Virus peptide core	Score	Pos in ADAMTS13	Virus	Protein	Position
LCGLASKPG	LAGLSTLPG	1.539	929	HCV	P27958	1772
VAVGPDVFQ	VVVSTDALM	1.147	83	HCV	P27958	1430
ILYCARAHG	IVYTACSHA	1.142	971	SARS-COV-2	P0DTD1	5624
LCGLASKPG	LSGMGTTAT	1.087	929	FLU-B	P03489	80
VAVGPDVFQ	LQVSSNVLD	0.810	83	HPV-B19	Q6TV13	4

Multiple Sclerosis: DRB1*15-associated autoimmune disease

DRB1*1501 allele is a very well-known susceptible allele for another autoimmune disease: multiple sclerosis (MS). MS is a neurodegenerative disease that is caused by attacks of the immune system to myelin basic protein (MBP), causing demyelination of neurons in the central nervous system. Infectious diseases are thought to be environmental triggers of MS. For example, it is known that antigens derived from *Mycobacterium avium* and *Escherichia coli* can induce MS-like disease in humanized mice (Harkiolaki et al., 2009). Hausmann et al. (1999) have demonstrated cross-reactivity of peptides from these bacteria with MBP. As we have PS-SCL data from DRB1*1501 allele, we wanted to test if it is possible to identify molecular mimicry using our approach.

M. avium and *E. coli* have a large proteome (6021 and 5062 proteins, respectively), and therefore, the number of potential T cell epitopes in the proteome of these bacteria is very large. However, very likely not all of these proteins are expressed at a high enough level to become immunogenic. To account for this, we created an ensemble of bacterial antigens, based on the presence of known epitopes from these bacteria in IEDB. In other words, we searched for all $CD4^+$, $CD8^+$ and B cell epitopes from *M. avium* and *E. coli* in IEDB and combined all source proteins from these epitopes as potential bacterial antigens, which resulted in 110 and 21 proteins for *M. avium* and *E. coli*, respectively. Human herpes virus (HHV-1) is used as control because of its similar size to the bacterial antigen ensemble and its non-existent relation with MS.

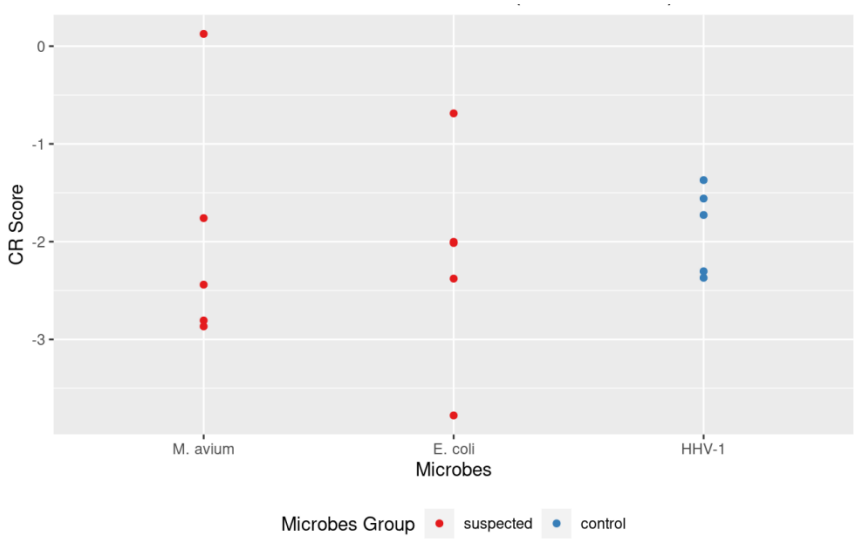


Figure 9. Cross-reactivity of MBP peptides with PS-SLC diversity model (DRB1*1501)

Figure 9 shows the peptides found with the best cross-reactivity score to MBP from *M. avium*, *E. coli* and HHV-1. The peptide with highest score is *M. avium* peptide VHFLRNVLA (mutator family transposase pos. 152-160) which is aligned with MBP peptide VHFFKNIVT (pos. 218-226, Table 4). This is a very exciting result because this exact peptide pair is proven experimentally to be cross-reactive (Hausmann et al., 1999). Of note this peptide alignment score is not as high as the

scores of peptides we identified as cross-reactive in the TTP analysis, but it stands out from the rest.

Table 4. Top 5 cross-reactivity score of MBP peptide using diversity-based prediction (DRB1*1501)

Self peptide core	Microbe peptide core	Score	Pos in MBP	Microbe	Protein	Position
VHFFKNIVT	VHFLRNVLA	0.126	217	M-avium	X8B370	152
LDVMASQKR	LDVMGSPIR	-0.687	128	E-coli	P0A6P6	443
LDVMASQKR	LVVWASAAR	-1.371	128	HHV-1	P36313	187
IGRFFGGDR	VMRFFGGLV	-1.559	172	HHV-1	P32888	240
LDVMASQKR	LDVINTNDV	-1.727	128	HHV-1	P10205	101

E. *coli* can also trigger MS in humanized mice (Harkiolaki et al., 2009), and an *E. coli* peptide that is cross-reactive with MBP is identified (Hausmann et al., 1999). This peptide, VHFISALHGSG (GTPase Der, pos. 347-355), is different than the one we predicted (LDVMGSPIR; GTPase Der, pos. 443-451). Moreover, this *E. coli* highest score in our result is much lower than that of *M. avium* (Fig. 9, Table 4). This is actually still in accordance with the literature since MS-like disease in humanized mice triggered by *E. coli* is considerably milder than the one triggered by *M. avium*, indicating that *E. coli*'s antigens might have lower cross-reactivity with MBP. Moreover, the *E. coli* cross-reactive peptide that is reported in the literature is synthesized and it is not yet known whether or not this peptide can be a result of natural process of antigen processing and presentation pathway.

Taken together, we are confident to conclude that our model can successfully predict molecular mimicry in MS, as our pipeline can identify the same *M. avium* and MBP peptide pair as in the experimental data.

Discussion and Conclusion

NetMHCIIpan is a powerful tool to predict peptide binding to MHC class II molecules. It is based on NNAlign machine learning framework and is reported to have better accuracy than several other tools (Zhao and Sher, 2018). This method can also be used for less common alleles because it contains a single universal network that can predict peptide binding affinities for all MHC molecules as long as the protein sequence is known. Moreover, the NetMHCIIpan version 4.0 that we used in this study incorporated eluted ligand data as a main determinator of peptide-MHC binding, as opposed to binding affinity data that is used in the previous versions. This approach is proven to be more accurate (Reynisson et al., 2020), simply because eluted ligand data not only contain binding affinity information, but also protein processing, transport, and presentation. Given this powerful computational tool, we believe that time is ripe for predicting molecular mimicry as a possible environmental factor of autoimmune diseases, and we aim to make a proof-of-principle for TTP in this study.

Earlier studies strongly suggested that influenza virus can trigger TTP via molecular mimicry (Sorvillo et al., 2013; Verbij et al., 2016; Hrdinova et al., 2018). Therefore, to determine which influenza peptide can cross-react with which ADAMTS13 peptide, we made TCR-HLA-peptide binding predictions for influenza and ADAMTS13 peptides. We tried several models to be able to determine cross-reactivity in the context of TCR recognition, as there are no available methods for this step yet. Our model 3 (Fig. 5) highlights difference in cross-reactivity between influenza and control viruses in one of the susceptible HLA alleles (HLA DRB1*0301, but not in DRB1*1101). An ADAMTS13 peptide that is known to bind HLA molecules and trigger $CD4^+$ T cell response, ILIRDTHSL (Sorvillo et al., 2013; Verbij et al., 2016), is predicted amongst the most cross-reactive peptides.

The PS-SCL data gave us more insights on the importance of the TCR contact residues. Amino acid substitutions in several positions are stricter than first thought (Wang and Reinherz, 2002). For example, P1 and P4 are almost as important as P2, P3, and P5, while P8 is not as important. Moreover, the positions affecting TCR specificity seem to differ between different HLA molecules. With all that being said, we still do not know much about pMHC-TCR interaction, and our insight now is based on very limited data.

The prediction result with the PS-SCL-based diversity model is better than the original model (compare Fig. 7 with Fig. 5). There are clearer score differences for some influenza peptide alignments in the susceptible allele (DRB1*1501) analysis and there is no difference in score for influenza and control viruses in the protective allele (DRB1*0404) analysis, indicating no enhanced cross-reactivity. For HLA DRB1*1501, the known ADAMTS13 peptide ILIRDTHSL that is aligned with an influenza C peptide comes out as the best scoring peptide alignment, while two other alignments (from influenza A and SARS-CoV-2) have clearly higher scores than control viruses. The two influenza peptides, VLIADAKGL (influenza C nucleoprotein pos. 189-197) and PVLYINVAD (influenza A H9N2 neuraminidase pos. 296-304), are strong candidates to cause molecular mimicry in TTP.

The results also suggest the possibility of cross-reactivity between ADAMTS13 and SARS-CoV-2. This novel coronavirus has high cross-reactivity score in the diversity-based model of DRB1*1501. SARS-CoV-2 peptide ILRLGSPLS (ORF9b protein pos. 41-49) is a strong cross-reactive peptide candidate because of its high score in the diversity-based model. This result supports several reports about TTP onset after SARS-CoV-2 infection and suggests molecular mimicry as the possible underlying mechanism of those TTP onset (Capecchi et al., 2020; Cohen et al., 2021).

Our results on an association between TTP and influenza viruses are further solidified by our multiple sclerosis (MS) analysis. Unlike TTP, molecular mimicry in MS is very well studied. Hausmann et al. (1999) found that peptides from *Mycobacterium avium* and *Escherichia coli* can stimulate myelin basic protein (MBP)-specific T cell clones and Harkiolaki et al. (2009) demonstrated that humanized mice injected with those peptides can induce MS-like disease. Our result shows that the exact MBP and *M. avium* peptide alignment, VHFFKNIVT / VHFLRNVLA, is the best-scoring peptide alignment and its score is significantly higher than that of controls, suggesting cross-reactivity (Fig. 8 and Table 4). This congeniality between our model prediction and the well-established experimental result is a proof of principle of our model.

The length of peptides that bind to MHC class II molecules vary greatly. Peptides of length 13-25 are considered to be ideal for MHC class II binding. We use length of 15 in this analysis based on the fact that the ADAMTS13 peptides that are known to bind MHC II and trigger $CD4^+$ T cell response are of length 15. We also repeated the analysis with length 16 and basically got the same result, so we assume a little length difference will not matter much. Even though it will be less likely for a peptide with considerably different lengths, for instance more than 20, to bind MHC class II molecules naturally, it may be interesting to perform a broader peptide length analysis in future studies.

Needless to say, we need MHC-binding and T cell response assays to confirm whether the peptides identified in this study can bind to MHC molecules and are cross-reactive with self antigens. These experiments could also determine where the threshold of our cross-reactivity score should be. In this analysis, to pick out which peptide pair will most likely to be cross-reactive is still very subjective, and the cross-reactivity score varies greatly between experiments. In MS analysis, the highest score is only 0.126 but the identified peptide has been proved to be cross-reactive, while in TTP analysis, the scores could be as high as 1.823 and there are a lot of peptides that score above 0.

Unfortunately, we could not find an influenza peptide that is cross-reactive with known ADAMTS13 peptide FINVAPHAR on HLA DRB1*1101 (and HLA DRB1*1501 in the diversity model). HLA DRB1*1101 is the most studied allele and is known as a predisposing allele of TTP. The CUB2 domain peptide FINVAPHAR is also well studied. It is found to bind to HLA DRB1*1101 molecule and induces $CD4^+$ T cell reaction. Therefore, it is disappointing that our model did not identify FINVAPHAR to be cross-reactive with any influenza peptides. It is possible that this

peptide will be identified as cross-reactive if we use PS-SCL data from HLA DRB1*1101 reactive T cell clones. Unfortunately, we did not find any study that has this data.

It could also possible that HLA DRB1*1101 is not the direct cause of TTP. The HLA DR gene has a strong linkage disequilibrium with HLA DQ (Sinkovits et al., 2017). The HLA DQ haplotypes could actually be the ones that are directly involved in molecular mimicry causing TTP. Sinkovits et al. (2017) found that several HLA DR alleles have higher odds on becoming a predisposing or protective factor of TTP when they are linked together with HLA DQ alleles. This could also explain why some studies have different HLA alleles as predisposing and protective alleles of TTP.

In conclusion, our model is able to demonstrate molecular mimicry mechanism between influenza and ADAMTS13 and predict some influenza peptides that may cause cross-reactivity. We build our model by taking into account amino acid diversity of each peptide position from T cell response PS-SCL data. Our prediction is supported by the fact that the ADAMTS13 peptide that is known to bind HLA molecules and trigger $CD4^+$ T cell response, ILIRDTHSL, is predicted amongst the most cross-reactive peptide, and further supported by the MS analysis result that is in accordance with experimental data.

References

Andreatta, M., Jurtz, V. I., Kaever, T., Sette, A., Peters, B., & Nielsen, M. (2017). Machine learning reveals a non-canonical mode of peptide binding to MHC class II molecules. *Immunology*, *152*(2), 255–264. https://doi.org/10.1111/imm.12763

Begon, M., Townsend, C. R., & Harper, J. L. (2006). *Ecology: From individuals to ecosystems*. Blackwell Pub.

Birnbaum, M. E., Mendoza, J. L., Sethi, D. K., Dong, S., Glanville, J., Dobbins, J., Özkan, E., Davis, M. M., Wucherpfennig, K. W., & Garcia, K. C. (2014). Deconstructing the peptide-MHC specificity of T cell recognition. *Cell*, *157*(5), 1073–1087. https://doi.org/10.1016/j.cell.2014.03.047

Capecchi, M., Mocellin, C., Abbruzzese, C., Mancini, I., Prati, D., & Peyvandi, F. (2020). Dramatic presentation of acquired Thombotic thrombocytopenic purpura associated with covid-19. *Haematologica*, *105*(10). https://doi.org/10.3324/haematol.2020.262345

Chicz, R. M., Urban, R. G., Lane, W. S., Gorga, J. C., Stern, L. J., Vignali, D. A., & Strominger, J. L. (1992). Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. *Nature*, *358*(6389), 764–768. https://doi.org/10.1038/358764a0

Cohen, M. K., Sheena, L., Shafir, Y., Yahalom, V., Gafter-Gvili, A., & Spectre, G. (2021). An early unexpected immune thrombotic thrombocytopenic purpura relapse associated with SARS-COV-2 infection: A case report and literature review. *Acta Haematologica*, *144*(6), 678–682. https://doi.org/10.1159/000514283

Coppo, P., Busson, M., Veyradier, A., Wynckel, A., Poullin, P., Azoulay, E., Galicier, L., & Loiseau, P. (2010). HLA-DRB1*11: A strong risk factor for acquired severe ADAMTS13 deficiency-related idiopathic thrombotic thrombocytopenic purpura in Caucasians. *Journal of Thrombosis and Haemostasis*, *8*(4), 856–859. https://doi.org/10.1111/j.1538-7836.2010.03772.x

Damian, R. T. (1964). Molecular mimicry: Antigen sharing by parasite and host and its consequences. *The American Naturalist*, *98*(900), 129–149. https://doi.org/10.1086/282313

Dias, P. J., & Gopal, S. (2009). Refractory thrombotic thrombocytopenic purpura following influenza vaccination. *Anaesthesia*, *64*(4), 444–446. https://doi.org/10.1111/j.1365-2044.2008.05823.x

Fujinami, R. S., & Oldstone, M. B. (1985). Amino acid homology between the encephalitogenic site of myelin basic protein and virus: Mechanism for autoimmunity. *Science*, *230*(4729), 1043–1045. https://doi.org/10.1126/science.2414848

Fujinami, R. S., Oldstone, M. B., Wroblewska, Z., Frankel, M. E., & Koprowski, H. (1983). Molecular mimicry in virus infection: Crossreaction of measles virus phosphoprotein or of herpes simplex virus protein with human intermediate filaments. *Proceedings of the National Academy of Sciences*, *80*(8), 2346–2350. https://doi.org/10.1073/pnas.80.8.2346

Gelder, C. M., Welsh, K. I., Faith, A., Lamb, J. R., & Askonas, B. A. (1995). Human CD4+ T-cell repertoire of responses to influenza A virus hemagglutinin after recent natural infection. *Journal of Virology*, *69*(12), 7497–7506. https://doi.org/10.1128/jvi.69.12.7497-7506.1995

Harkiolaki, M., Holmes, S. L., Svendsen, P., Gregersen, J. W., Jensen, L. T., McMahon, R., Friese, M. A., van Boxel, G., Etzensperger, R., Tzartos, J. S., Kranc, K., Sainsbury, S.,

Harlos, K., Mellins, E. D., Palace, J., Esiri, M. M., van der Merwe, P. A., Jones, E. Y., & Fugger, L. (2009). T cell-mediated autoimmune disease due to low-affinity crossreactivity to common microbial peptides. *Immunity*, *30*(3), 348–357. https://doi.org/10.1016/j.immuni.2009.01.009

Hausmann, S., Martin, M., Gauthier, L., & Wucherpfennig, K. W. (1999). Structural features of autoreactive TCR that determine the degree of degeneracy in peptide recognition. *Journal of immunology (Baltimore, Md. : 1950)*, *162*(1), 338–344.

Hermann, R., Pfeil, A., Busch, M., Kettner, C., Kretzschmar, D., Hansch, A., La Rosée, P., & Wolf, G. (2010). Schwerste thrombotisch-thrombozytopenische Purpura (TTP) nach H1N1-Vakzinierung [Very severe thrombotic thrombocytopenic purpura (TTP) after H1N1 vaccination]. *Medizinische Klinik (Munich, Germany : 1983)*, *105*(9), 663–668. https://doi.org/10.1007/s00063-010-1107-6

Hrdinová, J., Verbij, F. C., Kaijen, P. H. P., Hartholt, R. B., van Alphen, F., Lardy, N., ten Brinke, A., Vanhoorelbeke, K., Hindocha, P. J., De Groot, A. S., Meijer, A. B., Voorberg, J., & Peyron, I. (2018). Mass spectrometry-assisted identification of ADAMTS13-derived peptides presented on HLA-DR and HLA-DQ. *Haematologica*, *103*(6), 1083–1092. https://doi.org/10.3324/haematol.2017.179119

John, M.-L., Hitzler, W., & Scharrer, I. (2011). The role of human leukocyte antigens as predisposing and/or protective factors in patients with idiopathic thrombotic thrombocytopenic purpura. *Annals of Hematology*, *91*(4), 507–510. https://doi.org/10.1007/s00277-011-1384-z

Joly, B. S., Coppo, P., & Veyradier, A. (2017). Thrombotic thrombocytopenic purpura. *Blood*, *129*(21), 2836–2846. https://doi.org/10.1182/blood-2016-10-709857

Jonsson, M. K., Hammenfors, D., Oppegaard, O., Bruserud, Ø., & Kittang, A. O. (2015). A 35-year-old woman with influenza a-associated thrombotic thrombocytopenic purpura. *Blood Coagulation & Fibrinolysis*, *26*(4), 469–472. https://doi.org/10.1097/mbc.0000000000000255

Kaneko, H., Ohkawara, Y., Nomura, K., Horiike, S., & Taniwaki, M. (2004). Relapse of idiopathic thrombocytopenic purpura caused by influenza A virus infection: A case report. *Journal of Infection and Chemotherapy*, *10*(6), 364–366. https://doi.org/10.1007/s10156-004-0343-1

Kaplan, M. H., & Meyeserian, M. (1962). An immunological cross-reaction between group-a streptococcal cells and human heart tissue. *The Lancet*, *279*(7232), 706–710. https://doi.org/10.1016/s0140-6736(62)91653-7

Kim, Y., Sidney, J., Pinilla, C., Sette, A., & Peters, B. (2009). Derivation of an amino acid similarity matrix for peptide:MHC binding and its application as a Bayesian prior. *BMC Bioinformatics*, *10*(1). https://doi.org/10.1186/1471-2105-10-394

Koh, Y. R., Hwang, S.-H., Chang, C. L., Lee, E. Y., Son, H. C., & Kim, H. H. (2012). Thrombotic thrombocytopenic purpura triggered by influenza A virus subtype H1N1 infection. *Transfusion and Apheresis Science*, *46*(1), 25–28. https://doi.org/10.1016/j.transci.2011.10.024

Kosugi, N., Tsurutani, Y., Isonishi, A., Hori, Y., Matsumoto, M., & Fujimura, Y. (2010). Influenza A infection triggers thrombotic thrombocytopenic purpura by producing the anti-adamts13 IGG inhibitor. *Internal Medicine*, *49*(7), 689–693. https://doi.org/10.2169/internalmedicine.49.2957

Lee, C.-Y., Wu, M.-C., Chen, P.-Y., Chou, T.-Y., & Chan, Y.-J. (2011). Acute immune thrombocytopenic purpura in an adolescent with 2009 novel H1N1 influenza A virus

infection. *Journal of the Chinese Medical Association*, *74*(9), 425–427. https://doi.org/10.1016/j.jcma.2011.08.010

Levy, G. G., Nichols, W. C., Lian, E. C., Foroud, T., McClintick, J. N., McGee, B. M., Yang, A. Y., Siemieniak, D. R., Stark, K. R., Gruppo, R., Sarode, R., Shurin, S. B., Chandrasekaran, V., Stabler, S. P., Sabio, H., Bouhassira, E. E., Upshaw, J. D., Ginsburg, D., & Tsai, H.-M. (2001). Mutations in a member of the ADAMTS gene family cause thrombotic thrombocytopenic purpura. *Nature*, *413*(6855), 488–494. https://doi.org/10.1038/35097008

Li, F., Leier, A., Liu, Q., Wang, Y., Xiang, D., Akutsu, T., Webb, G. I., Smith, A. I., Marquez-Lago, T., Li, J., & Song, J. (2020). Procleave: Predicting protease-specific substrate cleavage sites by combining sequence and structural information. *Genomics, Proteomics & Bioinformatics*, *18*(1), 52–64. https://doi.org/10.1016/j.gpb.2019.08.002

Liu, J., & Gao, G. F. (2011). Major histocompatibility complex: Interaction with peptides. *ELS*. https://doi.org/10.1002/9780470015902.a0000922.pub2

Mamori, S., Amano, K., Kijima, H., Takagi, I., & Tajiri, H. (2008). Thrombocytopenic purpura after the administration of an influenza vaccine in a patient with autoimmune liver disease. *Digestion*, *77*(3-4), 159–160. https://doi.org/10.1159/000140977

Mariotte, E., Azoulay, E., Galicier, L., Rondeau, E., Zouiti, F., Boisseau, P., Poullin, P., de Maistre, E., Provôt, F., Delmas, Y., Perez, P., Benhamou, Y., Stepanian, A., Coppo, P., & Veyradier, A. (2016). Epidemiology and pathophysiology of adulthood-onset thrombotic microangiopathy with severe ADAMTS13 deficiency (thrombotic thrombocytopenic purpura): A cross-sectional analysis of the French National Registry for Thrombotic microangiopathy. *The Lancet Haematology*, *3*(5). https://doi.org/10.1016/s2352-3026(16)30018-7

Murphy, K., Travers, P., & Walport, M. (2008). *Janeway's Immunobiology*. Garland.

Reynisson, B., Barra, C., Kaabinejadian, S., Hildebrand, W. H., Peters, B., & Nielsen, M. (2020). Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data. *Journal of Proteome Research*, *19*(6), 2304–2315. https://doi.org/10.1021/acs.jproteome.9b00874

Rojas, M., Restrepo-Jiménez, P., Monsalve, D. M., Pacheco, Y., Acosta-Ampudia, Y., Ramírez-Santana, C., Leung, P. S. C., Ansari, A. A., Gershwin, M. E., & Anaya, J.-M. (2018). Molecular mimicry and autoimmunity. *Journal of Autoimmunity*, *95*, 100–123. https://doi.org/10.1016/j.jaut.2018.10.012

Sadler, J. E. (2015). What's new in the diagnosis and pathophysiology of thrombotic thrombocytopenic purpura. *Hematology*, *2015*(1), 631–636. https://doi.org/10.1182/asheducation-2015.1.631

Sakai, K., Kuwana, M., Tanaka, H., Hosomichi, K., Hasegawa, A., Uyama, H., Nishio, K., Omae, T., Hishizawa, M., Matsui, M., Iwato, K., Okamoto, A., Okuhiro, K., Yamashita, Y., Itoh, M., Kumekawa, H., Takezako, N., Kawano, N., Matsukawa, T., … Matsumoto, M. (2020). Hla loci predisposing to immune TTP in Japanese: Potential role of the shared ADAMTS13 peptide bound to different HLA-DR. *Blood*, *135*(26), 2413–2419. https://doi.org/10.1182/blood.2020005395

Schneider, E. L., & Craik, C. S. (2009). Positional scanning synthetic combinatorial libraries for substrate profiling. *Proteases and Cancer*, 59–78. https://doi.org/10.1007/978-1-60327-003-8_4

Scully, M., Brown, J., Patel, R., Mcdonald, V., Brown, C. J., & Machin, S. (2010). Human leukocyte antigen association in idiopathic thrombotic thrombocytopenic purpura:

Evidence for an immunogenetic link. *Journal of Thrombosis and Haemostasis*, *8*(2), 257–262. https://doi.org/10.1111/j.1538-7836.2009.03692.x

Sinkovits, G., Szilágyi, Á., Farkas, P., Inotai, D., Szilvási, A., Tordai, A., Rázsó, K., Réti, M., & Prohászka, Z. (2017). The role of human leukocyte antigen DRB1-DQB1 haplotypes in the susceptibility to acquired idiopathic thrombotic thrombocytopenic purpura. *Human Immunology*, *78*(2), 80–87. https://doi.org/10.1016/j.humimm.2016.11.005

Sorvillo, N., van Haren, S. D., Kaijen, P. H., ten Brinke, A., Fijnheer, R., Meijer, A. B., & Voorberg, J. (2013). Preferential HLA-DRB1*11–dependent presentation of cub2-derived peptides by ADAMTS13-pulsed dendritic cells. *Blood*, *121*(17), 3502–3510. https://doi.org/10.1182/blood-2012-09-456780

Verbij, F. C., Turksma, A. W., de Heij, F., Kaijen, P., Lardy, N., Fijnheer, R., Sorvillo, N., ten Brinke, A., & Voorberg, J. (2016). CD4+ T cells from patients with acquired thrombotic thrombocytopenic purpura recognize CUB2 domain-derived peptides. *Blood*, *127*(12), 1606–1609. https://doi.org/10.1182/blood-2015-10-668053

Wang, J.-huai, & Reinherz, E. L. (2002). Structural basis of T cell recognition of peptides bound to MHC molecules. *Molecular Immunology*, *38*(14), 1039–1049. https://doi.org/10.1016/s0161-5890(02)00033-0

Zhao, Y., Gran, B., Pinilla, C., Markovic-Plese, S., Hemmer, B., Tzou, A., Whitney, L. W., Biddison, W. E., Martin, R., & Simon, R. (2001). Combinatorial peptide libraries and biometric score matrices permit the quantitative analysis of specific and degenerate interactions between clonotypic TCR and MHC peptide ligands. *The Journal of Immunology*, *167*(4), 2130–2141. https://doi.org/10.4049/jimmunol.167.4.2130

Supplementary Data

Table 1a. Substitution matrix obtained from Zhao et al. (2001) (GP5F11; DRB1*0404)

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
A	1.06	5.44	0.67	1.21	0.74	3.97	10.75	1.49	2.25	2.21
C	1.21	1.19	0.82	1.56	1.42	1.65	2.26	1.28	1.53	1
D	0.88	3.35	1.02	1.14	0.94	1.01	1.9	1.54	1.18	0.94
E	0.57	1.01	0.74	1.23	0.97	1.19	0.8	0.93	1.28	1.27
F	18.98	3.14	1.02	1.6	0.82	1.21	5.41	1.29	10.86	9.43
G	0.91	0.86	0.84	1.65	1.23	4.65	29.95	8.48	2.5	1.19
H	2.09	2.18	1.62	4.47	5.44	3.43	7.88	1.38	4.73	1.52
I	1.26	16.35	1.25	9.52	1.22	11.1	2.6	0.94	3.12	10.94
K	1.35	1.79	71.66	4.09	1.35	0.7	1.39	30.65	7.02	1.08
L	1.64	18.74	1.01	8.74	1.77	1.81	2.83	1.27	5.9	14.05
M	1.31	28.4	1.18	2.23	2.26	2.32	2	3.26	9.74	12.22
N	1.02	1.17	0.69	1.09	19.57	12.27	2.88	2.17	3.44	1.08
P	1.11	0.86	0.92	3.1	0.9	1.25	15.16	3.88	3.73	1.04
Q	1.91	1.94	0.73	59.27	3.06	9.08	2.87	1.94	4.29	2.34
R	3.37	1.04	1.06	1.46	1.34	1.65	0.94	38.99	10.57	0.86
S	1.49	0.98	0.88	0.83	1.93	16.9	4.94	1.96	2.57	1.01
T	1.19	1.37	1.34	2.65	0.77	17.54	4.75	2.33	1.6	1.4
V	1.21	15.65	1.8	19.85	1.03	4.13	3.21	2.1	5.29	4.43
W	56.83	1.45	1.15	1.17	0.76	1.7	1.48	0.8	2.56	0.95
Y	30.63	2.32	1.1	8.5	1.05	0.86	4.29	1.33	9.02	3.77

Table 1b. Substitution matrix obtained from Birnbaum et al. (2014) (Ob.1A12 and Ob.2F3 (averaged); DRB1*1501)

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
A	0	0	0	0	0	0.245	0.025	0.07	0.11	0.1
C	0	0	0	0	0	0.005	0.01	0.01	0.095	0.025
D	0	0	0	0	0	0.015	0.005	0.02	0	0.02
E	0	0	0	0	0	0	0.06	0.04	0	0.02
F	0.01	0	1	0.46	0	0	0.015	0.005	0	0.03
G	0	0	0	0	0	0.23	0.02	0.06	0.045	0.03
H	0	1	0	0	0	0	0.09	0.05	0.005	0.03
I	0.45	0	0	0.02	0	0	0.04	0.02	0.06	0.02
K	0	0	0	0	0.715	0	0	0.04	0	0.045
L	0.19	0	0	0.1	0	0	0.17	0.105	0.245	0.08
M	0.03	0	0	0.02	0	0.045	0.015	0.04	0.035	0.04
N	0	0	0	0	0	0.175	0.095	0.02	0.005	0.02
P	0	0	0	0	0	0	0.095	0.1	0	0.03
Q	0	0	0	0	0	0.035	0.1	0.05	0.015	0.03
R	0	0	0	0	0.285	0	0.1	0.12	0	0.145
S	0	0	0	0	0	0.175	0.035	0.09	0.035	0.095
T	0	0	0	0	0	0.055	0.04	0.06	0.04	0.055
V	0.29	0	0	0.02	0	0	0.04	0.07	0.29	0.105
W	0.01	0	0	0.08	0	0	0.03	0.005	0	0.05
Y	0.01	0	0	0.3	0	0	0.01	0.015	0	0.03