



OPEN ACCESS

EDITED BY

Sheng Liu,
Indiana University Bloomington, United States

REVIEWED BY

Bin Hu,
Los Alamos National Laboratory (DOE),
United States
Fu Gao,
Yale University, United States
Qingxia Yao,
University of Alberta, Canada

*CORRESPONDENCE

Xavier Daura,
✉ xavier.daura@uab.cat

RECEIVED 29 September 2024

ACCEPTED 25 November 2024

PUBLISHED 10 December 2024

CITATION

Guerra F, Junet V, Farrés J and Daura X (2024)
MMPred: a tool to predict peptide mimicry
events in MHC class II recognition.
Front. Genet. 15:1500684.
doi: 10.3389/fgene.2024.1500684

COPYRIGHT

© 2024 Guerra, Junet, Farrés and Daura. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

MMPred: a tool to predict peptide mimicry events in MHC class II recognition

Filippo Guerra^{1,2}, Valentin Junet^{1,2}, Judith Farrés¹ and
Xavier Daura^{2,3,4*}

¹Anaxomics Biotech, Barcelona, Spain, ²Institute of Biotechnology and Biomedicine, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain, ³Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain, ⁴Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III, Cerdanyola del Vallès, Spain

We present MMPred, a software tool that integrates epitope prediction and sequence alignment algorithms to streamline the computational analysis of molecular mimicry events in autoimmune diseases. Starting with two protein or peptide sets (e.g., from human and SARS-CoV-2), MMPred facilitates the generation, investigation, and testing of mimicry hypotheses by providing epitope predictions specifically for MHC class II alleles, which are frequently implicated in autoimmunity. However, the tool is easily extendable to MHC class I predictions by incorporating pre-trained models from CNN-PepPred and NetMHCpan. To evaluate MMPred's ability to produce biologically meaningful insights, we conducted a comprehensive assessment involving *i*) predicting associations between known HLA class II human autoepitopes and microbial-peptide mimicry, *ii*) interpreting these predictions within a systems biology framework to identify potential functional links between the predicted autoantigens and pathophysiological pathways related to autoimmune diseases, and *iii*) analyzing illustrative cases in the context of SARS-CoV-2 infection and autoimmunity. MMPred code and user guide are made freely available at <https://github.com/ComputBiol-IBB/MMPRED>.

KEYWORDS

molecular mimicry, epitope prediction, sequence alignment, MHC class II, autoimmune disease, SARS-CoV-2

1 Introduction

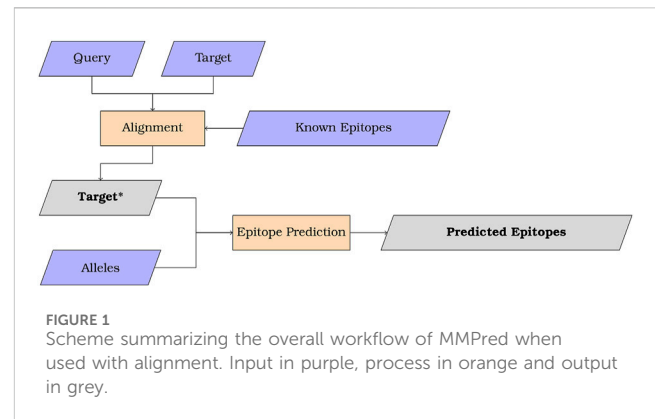
Epidemiological, clinical, and experimental evidence supports the association between infections and autoimmune diseases, with molecular mimicry proposed as one of the key mechanisms underlying this relationship (Oldstone, 1998; Rojas et al., 2018). Molecular mimicry occurs when a pathogen-derived peptide shares sequence similarity with a host peptide, and it is considered a strategy employed by pathogens to evade the immune system. By mimicking self-protein sequence fragments, pathogens exploit the immune system's tolerance for these molecules, effectively avoiding detection. This mimicry can mislead the immune response, activating autoreactive T-cells and/or producing cross-reactive antibodies that target both the pathogen's peptides and the host's own tissues. As a result, the immune system may inadvertently attack the body, contributing to the onset of autoimmune disorders. This mechanism is known to play a role in diseases such as Guillain-Barré syndrome (Maguire et al., 2024) or rheumatic fever (Cunningham, 2014), where infections by specific viruses or bacteria, respectively, are linked to autoimmunity.

The identification of molecular mimicry events is challenging due to the vast number of potential pathogen-derived peptides and the limited number of known autoantigens.

The identification or prediction of peptide mimicry events could therefore serve as a valuable clinical tool. From a theoretical standpoint, predicting peptide mimicry involves identifying similarities between self and exogenous proteins (Doxey and McConkey, 2013) and predicting their recognition by the immune system (Rojas et al., 2018). In this context, bioinformatics can provide valuable insights through the application of sequence alignment and epitope-prediction algorithms, particularly for Major Histocompatibility Complex (MHC) epitopes (Caron et al., 2015; Wang et al., 2010). To our knowledge, CRESSP (An et al., 2022) is the only tool currently integrating and streamlining both sequence alignment and epitope prediction. However, CRESSP primarily focuses on B-cell epitopes. Although the SARS-CoV-2 study by An et al. (2022) included MHC class II epitopes as well, the actual tool (<https://pypi.org/project/cressp/>) only incorporates MHC class I epitope predictions, as provided by MHCflurry (O'Donnell et al., 2020).

Here, we present the program MMPred (Molecular Mimicry Predictor), a tool that integrates sequence alignment and MHC class II epitope-prediction algorithms into a single pipeline. The tool is designed to be flexible, user-friendly and amenable to non-expert use, requiring as sole inputs the fasta files for the exogenous (query) and endogenous (target) protein (or peptide) sets and a list of MHC alleles of interest to perform the predictions for. The application of sequence alignment is optional, and if used, the two sets of sequences are compared and epitope prediction is applied to those target sequences that show a significant alignment with query sequences. The tool is an extension of CNN-PepPred (Junet and Daura, 2021), and offers the possibility to include predictions from NetMHCIIpan4.1, both with the BA (trained on binding affinities) and EL (trained on eluted ligands) models (Reynisson et al., 2020). The tool is programmed in such a way that additional predictors, including MHC class I epitope predictors, can be added with minimal effort. The alignments can be performed with BLASTp (Altschul et al., 1990) or by means of a position-specific scoring matrix (PSSM) with PSI-BLAST (Schäffer et al., 2001).

We have evaluated the capacity of MMPred to produce biologically meaningful results by *i*) predicting the association of known HLA class II human autoepitopes to microbial-peptide mimicry, *ii*) evaluating and interpreting these predictions in a systems biology framework, where the predicted autoantigens were tested for possible functional relations to pathophysiological pathways associated to autoimmune disease, and *iii*) analyzing example cases in the context of autoimmunity and SARS-CoV-2 infection (Ehrenfeld et al., 2020; Knight et al., 2021). We note that a statistical evaluation of the performance of the tool would not be meaningful in this case since *i*) the number of human peptides that are known to be (and not be) autoreactive as a result of microbial-peptide mimicry is very small and *ii*) we would be mostly evaluating the prediction performance of CNN-PepPred and NetMHCIIpan, which has been already done and is not the purpose of this study. For the alignment strategy, both BLASTp and PSI-BLAST produced biologically relevant matches. Although PSI-BLAST yielded higher average scores in the functional evaluation, it significantly reduced the number of hits at the same significance threshold. Therefore, we recommend using both



alignment methods for a more comprehensive analysis. We discuss some of the predicted autoantigens in further detail, with helicase MOV-10 standing as one of the most interesting cases.

2 Methods

2.1 MMPred

This study introduces MMPred, a software tool integrating epitope prediction and sequence alignment algorithms to simplify the setup of computational analyses aimed at the generation, investigation or testing of hypotheses relative to molecular mimicry events in the context of autoimmune diseases. As it stands, the tool provides epitope predictions for MHC class II only, as alleles involved in autoimmunity belong often to this class (Fiorillo et al., 2017). But the tool can be easily extended to MHC Class I prediction by incorporating the corresponding pre-trained models from CNN-PepPred and NetMHCpan (Reynisson et al., 2020).

2.2 Algorithm

The MMPred algorithm is illustrated in Figure 1 for uses combining sequence alignment and epitope prediction. In addition, the program may be also used without the alignment feature to streamline epitope prediction using NetMHCIIpan and CNN-PepPred.

2.2.1 Software specifications

The software has been developed using Python3.6.8 in a Linux environment. The Python3 libraries used are Pandas (<https://pandas.pydata.org/>), NumPy (Harris et al., 2020), Matplotlib (Hunter, 2007), pickle (<https://github.com/python/cpython/blob/3.6/Lib/pickle.py>), sklearn (Pedregosa et al., 2011) and tensorflow (Abadi et al., 2016). Software, installation instructions and program user-guide are available in GitHub (<https://github.com/ComputBiol-IBB/MMPRED>).

2.2.2 Input

The program takes a set of protein sequences in the form of a fasta file (QUERY) and a list of MHC alleles (ALLELES) as minimal

TABLE 1 Results of the supervised evaluation.

Human autoepitope	AID	Microbial epitope	Organism	Same allele
H1-1 (62-70)	MS	PPE68 (63-68)	MT	-
H1-2 (59-67)	MS	PPE69 (63-68)	MT	-
H1-2 (203-211)	MS	HbhA (167-185)	MT	DRB5*01:01
H1-4 (210-218)	MS	HbhA (168-185)	MT	DRB5*01:01
H1-4 (210-218)	MS	RplV (152-157)	MT	DRB5*01:01
H2BC3 (66-74)	MS	EsxB (18-24)	MT	-
MPO (234-242)	MS	Spike (326-330)	SARS-CoV-2	DRB1*15:01
IFT57 (60-68)	MS	Spike (457-462)	SARS-CoV-2	-
PTPRJ (693-701)	MS	Spike (778-785)	SARS-CoV-2	-
RPL31 (103-116)	MS	Spike (1065-1069)	SARS-CoV-2	-
RPL7A (60-68)	MS	Spike (1210-1218)	SARS-CoV-2	-
ARID4B (540-552)	MS	Replicase polyprotein 1a (972-977)	SARS-CoV-2	-
CCDC97 (248-256)	MS	Replicase polyprotein 1a (972-977)	SARS-CoV-2	-
GLT8D1 (48-56)	MS	Replicase polyprotein 1a (1910-1917)	SARS-CoV-2	-
TGFBI (235-244)	MS	Replicase polyprotein 1a (2147-2155)	SARS-CoV-2	-
VIM (52-64)	MS	Replicase polyprotein 1a (3381-3389)	SARS-CoV-2	-
HLA-A (57-65)	MS	Replicase polyprotein 1a (3991-3999)	SARS-CoV-2	-
HLA-A (57-66)	RA	Replicase polyprotein 1a (3991-3999)	SARS-CoV-2	-
MRPS15 (196-204)	MS	Nucleoprotein (86-93)	SARS-CoV-2	-
PLXDC2 (38-49)	MS	Nucleoprotein (267-274)	SARS-CoV-2	-
ACTA2 (153-161)	MS	ORF3a protein (164-176)	SARS-CoV-2	-
H3-4 (21-29)	MS	ORF8 protein (51-56)	SARS-CoV-2	-
RPL5 (18-26)	MS	Tegument protein UL46 (486-490)	HHV-1	-
JAK2 (102-110)	MS	gE (38-42)	HHV-3	-

Human autoepitope: human protein name and position in the sequence of the predicted autoepitope; AID: autoimmune disease associated to the microorganism (MS: multiple sclerosis, RA: rheumatoid arthritis); Microbial epitope: microbial protein name and position in the sequence of the known microbial epitope; Organism: corresponding microorganism (MT: *Mycobacterium tuberculosis*, HHV: Human Alphaherpesvirus); Same allele: predicted allele, only shown when the same allele is known to recognise both human and microbial epitopes at the experimental level. An extended version of this table is provided in the Supplementary Material file MEDS_vs_HADS.xlsx.

input. If no additional input is provided, the program runs an epitope prediction for the sequences in QUERY and each of the MHC alleles specified.

The user can specify if QUERY contains entire protein sequences (*protein* mode) or peptides (*peptide* mode). When in *protein* mode, the program makes a prediction for each fragment of size W in the protein (with a 1-residue step), where W is a user-defined parameter. Instead, when using the *peptide* mode a single prediction is performed for the full length of the peptide. We suggest to use the *peptide* mode only when QUERY contains short-length peptides, e.g., ≤ 25 residues.

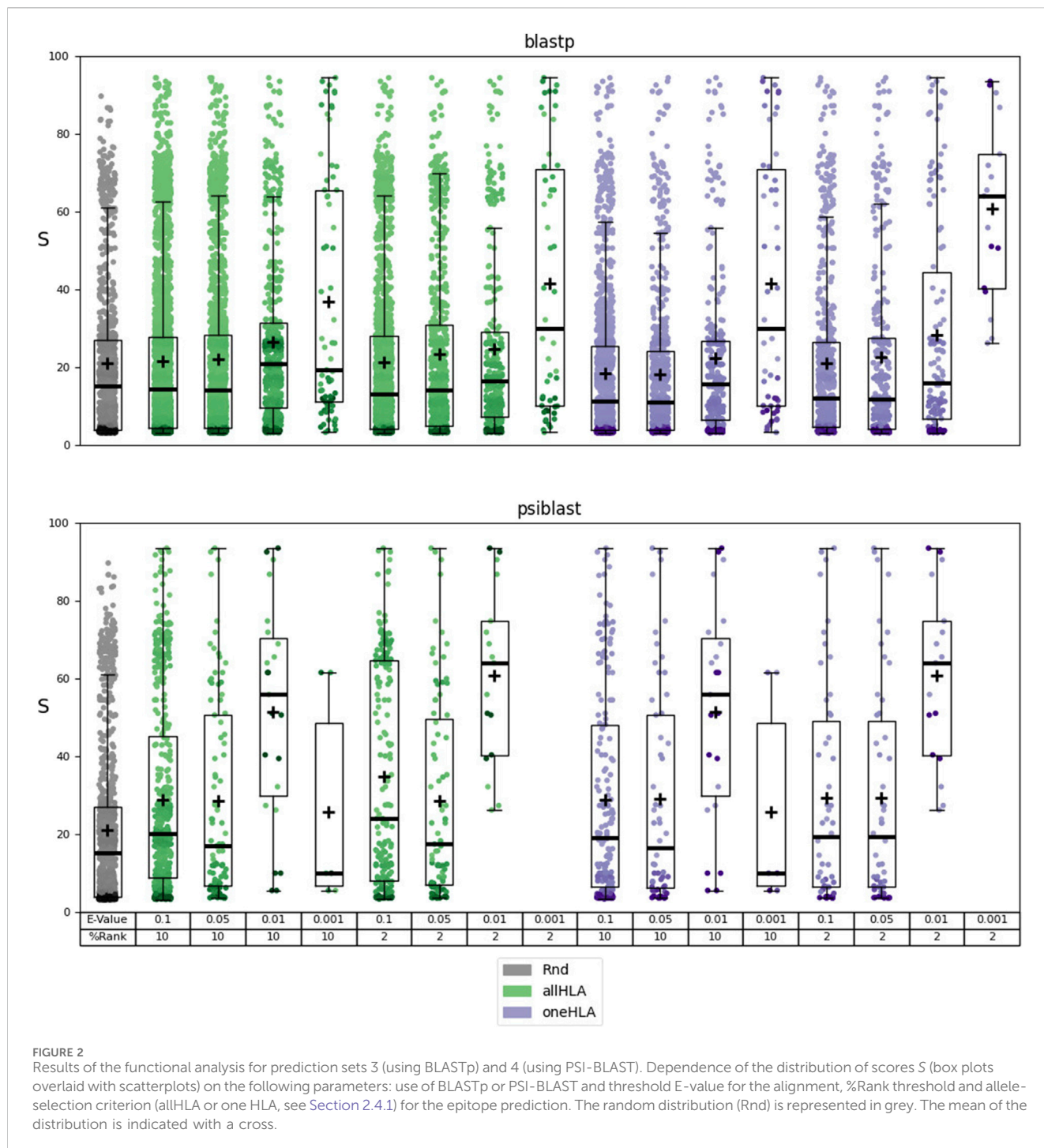
The user can provide a second fasta file (TARGET). In this case, an alignment is performed between the peptides in QUERY (if in *peptide* mode) or derived from QUERY (if in *protein* mode) and the TARGET sequences, to produce the TARGET* set (see Figure 1). TARGET* contains the sequences from TARGET that show significant (to a value defined by the user) alignment with

QUERY peptides. Epitope prediction is then applied to the TARGET* sequences.

In this study, we are using microbial sequences as QUERY and human ones as TARGET. Yet, QUERY and TARGET may be whatever the user thinks appropriate for the analysis in question.

2.2.3 Alignment

Protein sequence alignment is performed to identify the potential similarity between QUERY and TARGET sequences, if TARGET is specified. The alignment can be done using either of two strategies: *i*) using BLASTp to perform an ungapped alignment, with automatic adjustment of parameters for short input sequences or *ii*) using PSI-BLAST to compute a PSSM by aligning the sequences in QUERY against a user-defined set of epitopes provided in a fasta file, and then using the PSSM to perform a search in TARGET. This user-defined epitope set is identified in Figure 1 as *Known Epitopes*, and an example set used in this study is described in Section 2.3.5.



If the alignment satisfies a certain significance threshold (either E-value or bit score; by default E-Value < 0.05) the aligned TARGET sequence is stored in TARGET* for epitope prediction. The parameter (E-value or bit score) and its value can be defined by the user.

To perform the epitope prediction, a sequence of length $\geq W$ that we shall call “prediction window” has to be extracted from the alignment. If the length of the alignment is $< W$, then a prediction window of size W centered on the aligned sequence is extracted. If the number of extra residues at left and right cannot be the

same, the program automatically takes the extra residue to the right. If the window falls outside the ends of the sequence, the algorithm will take the first or last W residues accordingly. W is by default 15.

2.2.4 Epitope prediction

Whether the alignment is performed or not, epitope predictions are run using NetMHCIIpan 4.1, with both the model trained on Binding Affinity (BA) and that trained on mass spectrometry Eluted Ligands (EL) (Reynisson et al., 2020), and CNN-PepPred (Junet and

TABLE 2 Results of the functional analysis for prediction sets 3 (using BLASTp) and 4 (using PSI-BLAST).

Alignment	Alleles	%Rank	E-Value	N	Score ratio	p-value	Significance
BLASTp	allHLA	10	0.1	3144	1.11	3.41e-02	-
BLASTp	allHLA	10	0.05	1770	1.15	2.33e-02	-
BLASTp	allHLA	10	0.01	516	1.36	1.10e-13	***
BLASTp	allHLA	10	0.001	67	1.91	2.05e-06	***
BLASTp	allHLA	2	0.1	1594	1.1	1.99e-01	-
BLASTp	allHLA	2	0.05	794	1.2	1.07e-02	-
BLASTp	allHLA	2	0.01	301	1.28	4.25e-04	**
BLASTp	allHLA	2	0.001	50	2.14	9.51e-07	***
BLASTp	oneHLA	10	0.1	1346	0.96	9.88e-01	-
BLASTp	oneHLA	10	0.05	711	0.95	9.82e-01	-
BLASTp	oneHLA	10	0.01	299	1.15	1.26e-02	-
BLASTp	oneHLA	10	0.001	50	2.14	9.51e-07	***
BLASTp	oneHLA	2	0.1	618	1.09	3.53e-01	-
BLASTp	oneHLA	2	0.05	330	1.16	2.95e-01	-
BLASTp	oneHLA	2	0.01	134	1.47	2.43e-03	*
BLASTp	oneHLA	2	0.001	17	3.14	9.69e-10	***
PSI-BLAST	allHLA	10	0.1	487	1.49	1.53e-13	***
PSI-BLAST	allHLA	10	0.05	89	1.47	6.60e-04	**
PSI-BLAST	allHLA	10	0.01	23	2.66	3.85e-08	***
PSI-BLAST	allHLA	10	0.001	6	1.33	3.26e-01	-
PSI-BLAST	allHLA	2	0.1	184	1.8	3.03e-11	***
PSI-BLAST	allHLA	2	0.05	83	1.48	6.45e-04	**
PSI-BLAST	allHLA	2	0.01	17	3.14	9.69e-10	***
PSI-BLAST	allHLA	2	0.001	0	-	-	-
PSI-BLAST	oneHLA	10	0.1	188	1.49	1.22e-05	***
PSI-BLAST	oneHLA	10	0.05	56	1.5	1.04e-02	-
PSI-BLAST	oneHLA	10	0.01	23	2.66	3.85e-08	***
PSI-BLAST	oneHLA	10	0.001	6	1.33	3.26e-01	-
PSI-BLAST	oneHLA	2	0.1	50	1.52	1.08e-02	-
PSI-BLAST	oneHLA	2	0.05	50	1.52	1.08e-02	-
PSI-BLAST	oneHLA	2	0.01	17	3.14	9.69e-10	***
PSI-BLAST	oneHLA	2	0.001	0	-	-	-

Alleles: allele-selection criterion (see Section 2.4.1); Score ratio: mean-score ratio; N: size of the sample; p-value: from the Mann-Whitney U test; Significance: - (not significant), * (p-value < 0.01), ** (p-value < 0.001), *** (p-value < 0.0001).

Daura, 2021). The three predictions are kept and no consensus score is generated. Note that NetMHCIIpan reports the prediction score and the %Rank for each peptide-MHC pair. The %Rank is a normalized prediction score that enables comparison between different MHC alleles and models (BA and EL). The %Rank of a query sequence is determined by comparing the prediction score to a score distribution for a random set of natural peptides, with %

Rank = 1 meaning that the queried sequence obtained a prediction score in the highest 1% of the distribution. On the other hand, CNN-PepPred reports only prediction scores. To make results from CNN-PepPred comparable to NetMhcIIpan, a score distribution for natural peptides was generated for each MHC allele available in CNN-PepPred, using a random sample of 10,000 peptides extracted from UniRef50 (The UniProt Consortium, 2023).

TABLE 3 MMPred-predicted autoantigens from similarity to SARS-CoV-2 sequences.

Autoantigen	Epitope	SARS-CoV-2 protein	Aligned sequence	N positive
MYT1L	EEGDREEEE (125–133)	NSP3	DEDEEEGDCEEEE (110–122)	1
BAZ1A	VDGDEEEGQSEEEE (1229–1242)	NSP3	DEDEEEGDCEEEFE (110–124)	2
CHD5	DDDDDEEEGGCEEEED (398–413)	NSP3	DEDEEEGDCEEEE (110–122)	2
MCM8	YNYEPLTQL (199–207)	NSP5	YNYEPLTQ (237–244)	1
ATF7	FVCNAPGCG (7–15)	NSP13	PYVCNAPGC (47–55)	1
HELZ2	FTVIQPPG (2169–2177)	NSP13	QKYSTLQGGPGTGKS (275–289)	11
MOV10	KRFNVAVTRAKAL (903–915)	NSP13	NVNRFNVAITRAK (557–569)	5
MOV10L1	RFNVAITRPKAL (1131–1142)	NSP13	NVNRFNVAITRAK (557–569)	5
DNA2	LNVAITRAKH (1000–1009)	NSP13	RFNVAITRAK (560–569)	4
BRI3	VTRYPANSI (64–72)	NSP14	VDRYPANSIV (389–398)	7
PARVG	LHLLVALAKRFQ (140–151)	NSP15	LHLLIGLAKRF (248–258)	8
CALD1	VMSLKNQI (225–233)	NSP16	TAVMSLKEGQI (257–267)	2
MICAL3	YKKDKKKKA (1747–1755)	N	KKDKKKKADE (369–378)	3
SLC35E4	SVLYNLASF (265–273)	spike(BA.1-like)	SVLYNLASF (366–375)	1
UNC50	YKYLRLRFR (32–40)	spike(BA.4-like)	YNYLRLRFR (447–455)	7

N positive: number of positive predictions for the epitope (pairs of prediction method and allele).

2.2.5 Output

The output depends on the input parameters as follows:

- *No Alignment, Protein mode*: The program will return the predicted binding core (a sequence of length 9 for MHC class II binders (Wu et al., 2021)) for every window of size W achieving $\%Rank \leq 10$ in every protein included in QUERY, together with the ID of the protein, start and end position of the W -residue peptide, start and end position of the predicted core, prediction method, score, $\%Rank$ and the allele for which the prediction has been made.
- *No Alignment, Peptide mode*: The program will return the predicted 9-mer core for every peptide included in QUERY achieving $\%Rank \leq 10$, together with the ID of the peptide, start and end position of the predicted core, prediction method, score, $\%Rank$ and the allele for which the prediction has been made.
- *Alignment*: For every sequence in TARGET* achieving $\%Rank \leq 10$, the program will return the predicted 9-mer core, the TARGET sequence ID, the start and end position of the alignment in the TARGET sequence, the start and end position of the window of size W extracted from the alignment in the TARGET sequence, the start and end position of the predicted core in the TARGET sequence, identity, E-value and bit score of the alignment, the aligned TARGET sequence, the QUERY sequence ID, the start and end position in the QUERY sequence of the alignment, the aligned QUERY sequence, prediction method, score, $\%Rank$ and the allele for which the prediction has been made.

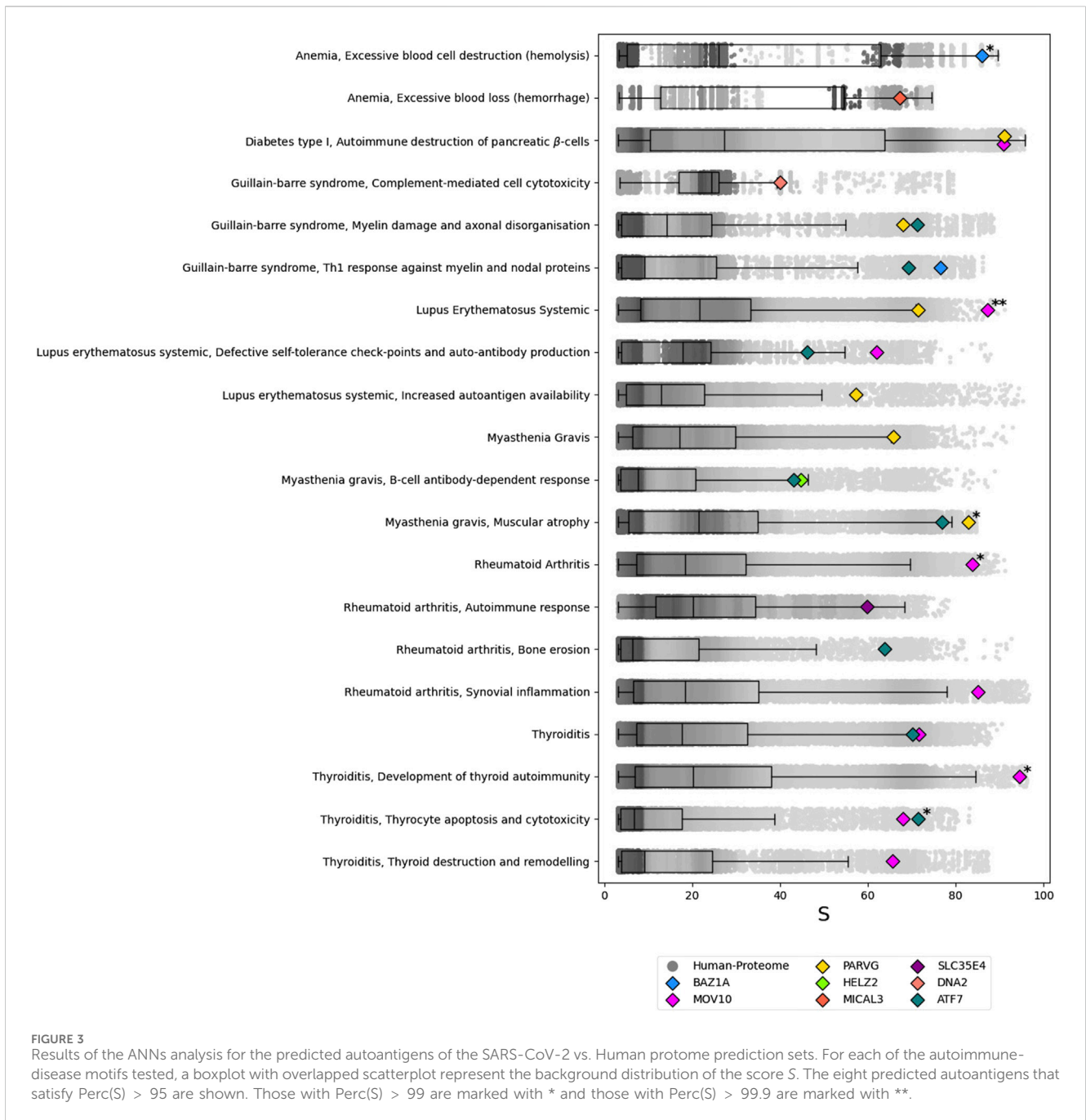
2.3 Evaluation datasets

2.3.1 Microbial epitope dataset (MEDS)

A dataset of known HLA class II epitopes from infectious agents was manually downloaded from IEDB (Vita et al., 2019) (<https://www.iedb.org>). The search terms “Epitope: Linear peptide”, “Epitope source: Bacteria, Virus”, “Host: Human”, “Assay: MHC ligand”, “Outcome: positive”, “MHC Restriction: Class II”, “Disease: any” were used (date 11/10/2023).

Epitopes lacking either the UniProtKB accession number (AC) (The UniProt Consortium, 2023) (<https://www.uniprot.org>) of the protein, start or end position in the sequence or the HLA allele for which they were tested, were excluded. Epitopes containing modified amino-acid residues were also discarded. Additionally, only epitopes from microorganisms known to be related to autoimmune diseases were considered. The relationship between an organism and the occurrence of autoimmune diseases was determined by review of the literature when not annotated in IEDB. For each organism with epitopes fulfilling the previous selection criteria, a PubMed (<https://pubmed.ncbi.nlm.nih.gov/>) search for reviews from the last 10 years using the keywords “organism_name AND (autoimmunity OR autoimmune)” was performed. As last filter, epitopes associated to HLA alleles with no model available in CNN-PepPred and NetMHCIIpan were also discarded.

After these filters, MEDS contained 3,676 epitopes from 88 proteins of 13 microorganisms, associated to 50 HLA class II alleles, with a total of 9,229 epitope-allele pairs (Supplementary Material Table 1 file MEDS_summary.xlsx).



2.3.2 Human autoepitope DataSet (HADS)

A dataset containing known human autoepitopes was downloaded from IEDB (11/10/2023). The search terms were: “Epitope: Linear peptide”, “Epitope source: Human”, “Host: Human”, “Assay: MHC ligand”, “Outcome: positive”, “MHC Restriction: Class II”, “Disease: autoimmune”.

To avoid redundancy in the form of nested sets, epitopes from the same protein that showed overlap and were linked to the same allele were merged into a single epitope. After the merging, only sequences with length ≥ 15 were kept. Epitopes associated to HLA alleles with no model available in CNN-PepPred and NetMHCIIpan were also discarded.

HADS thus contained 807 epitopes from 608 different human proteins, associated to 5 different HLA class II alleles in the context of Rheumatoid Arthritis and Multiple Sclerosis (Supplementary Material Table 2 file HADS_summary.xlsx).

2.3.3 Human proteome dataset (HPDS)

HPDS contains the sequences of the 20,426 reviewed human proteins found in UniProt (The UniProt Consortium, 2023) on the date of the download (18/10/2023).

2.3.4 SARS-CoV-2 proteome dataset (SC2DS)

SC2DS was generated from the SARS-CoV-2 reference proteome (Wu et al., 2020) (UniProt identifier: UP000464024,

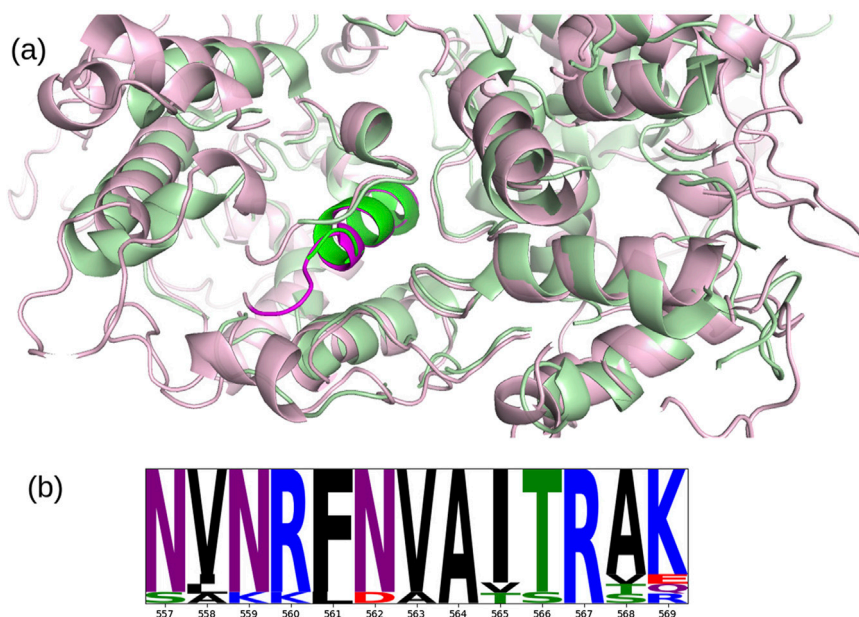


FIGURE 4
(A) Superposition of the crystallographic structure of NSP13 (PDB entry 6ZSL, chain B) (green) with the predicted AlphaFold structure of MOV10 (UniProt entry Q9HCE1) (pink), the highlighted α -helices in the center correspond to residues 557–569 of NSP13 and 901–913 of MOV10. Superposition and image were generated with PyMOL (<https://www.pymol.org/>). **(B)** Logo plot of the multiple sequence alignment for the NSP13 epitope. To facilitate visualization, a pseudocount of 0.1 is used and a min-max normalization of each position is applied. Image generated with the logomaker python package (Tareen and Kinney, 2020).

downloaded 11/01/2024) and the Pangolin variants (Rambaut et al., 2020) (downloaded 01/03/2024) A.23.1-like, A.23.1-like + E484K, Alpha_(B.1.1.7-like), AV.1-like, B.1.1.318-like, B.1.1.7-like + E484K, B.1.617.1-like, B.1.617.3-like, Beta_(B.1.351-like), Delta_(AY.4.2-like), Delta_(AY.4-like), Delta_(B.1.617.2-like), Delta_(B.1.617.2-like)+K417N, Epsilon_(B.1.427-like), Epsilon_(B.1.429-like), Eta_(B.1.525-like), Gamma_(P.1-like), Iota_(B.1.526-like), Lambda_(C.37-like), Mu_(B.1.621-like), Omicron_(BA.1-like), Omicron_(BA.2-like), Omicron_(BA.3-like), Omicron_(BA.4-like), Omicron_(BA.5-like), Omicron_(Unassigned), Omicron_(XBB.1.16-like), Omicron_(XBB.1.5-like), Omicron_(XBB.1-like), Omicron_(XBB-like), Theta_(P.3-like), XBB-parent1, XBB-parent2, XE-parent1, XE-parent2, Zeta_(P.2-like). The downloaded sequences were manually split into all possible overlapping fragments of size 15. Fragments of variant sequences were only kept if they had indels or mutations relative to the reference proteome. SC2DS thus contained 9,608 15-mers, of which 243 from Pangolin variants.

2.3.5 MHC class II epitopes dataset (MHCII-EDS)

MHCII-EDS contains all MHC class II epitopes available in IEDB (date 18/03/2024), obtained with the search terms: “Epitope: Linear peptide”, “Epitope source: Any”, “Host: Any”, “Assay: MHC ligand”, “Outcome: positive”, “MHC Restriction: Class II”, “Disease: Any”. A total of 485,020 epitopes were downloaded. Redundancy was eliminated by clustering the sequences at 95% identity with CD-HIT (Li and Godzik, 2006), then using the centroid as cluster representative to obtain a final set of 155,923 epitopes. This dataset was used for the computation of the PSSM in all the analyses where PSI-BLAST was used as alignment algorithm.

2.4 MMPred evaluation

2.4.1 Evaluation setup

To evaluate the algorithm, five sets of predictions were obtained:

1. Sequences from the human autoepitope dataset (HADS) (TARGET) that significantly align with sequences from the microbial epitope dataset (MEDS) (QUERY) using BLASTp and were positive for binding to HLA class II (with or without allele match, see below) with CNN-PepPred and/or NetMHCIIpan.
2. Same as prediction set 1 but using PSI-BLAST for the alignment.
3. Same as prediction set 1 but with the human proteome dataset (HPDS) as TARGET.
4. Same as prediction set 3 but using PSI-BLAST for the alignment.
5. Sequences from HPDS (TARGET) that significantly align with sequences from the SARS-CoV-2 proteome dataset (SC2DS) (QUERY) using BLASTp and were positive for binding to HLA class II with CNN-PepPred and/or NetMHCIIpan.
6. Same as prediction set 5 but using PSI-BLAST for the alignment.

Prediction sets 1 and 2 were based on a threshold E-value of 0.05 for the alignments and a %Rank ≤ 2 as condition for binding—Reynisson et al. (2020) defined peptides with %Rank ≤ 2 as strong binders and peptides with $2 < \text{\%Rank} \leq 10$ as weak binders. Prediction sets 3 and 4 were instead obtained in replicates by using threshold E-values from 0.1 to 0.001 and threshold %Rank

values of 2 and 10. Owing to the results obtained for prediction sets 3 and 4, prediction sets 5 and 6 were based on a threshold E-value of 0.01 for the alignments and a %Rank ≤ 2 .

Prediction sets 1 to 4 were evaluated according to two allele-selection criteria:

1. AllHLA: epitope prediction for HADS (or HPDS) sequences that had at least one significant alignment with MEDS sequences was performed for all alleles.
2. OneHLA: epitope prediction for HADS (or HPDS) sequences that had at least one significant alignment with MEDS sequences was only performed for the allele(s) corresponding to the microbial epitope-allele pair(s) indicated in MEDS.

Furthermore, they were also evaluated both ignoring and considering allele matches:

1. *Epitope* prediction: sequences in HADS (or HPDS) were considered to be predicted as mimicry-induced autoepitopes if there was at least one alignment with MEDS sequences that satisfied the threshold E-value and there was at least one prediction from CNN-PepPred or NetMHCIIpan (BA or EL models) that satisfied the threshold %Rank for any HLA allele.
2. *Epitope-allele* prediction: epitope-allele pairs in HADS (or HPDS) were considered to be predicted as autoepitope-allele pairs if there was at least one alignment with MEDS that satisfied the threshold E-value, and there was at least one prediction from CNN-PepPred or NetMHCIIpan (BA or EL models) that satisfied the threshold %Rank for the same HLA allele of the MEDS pair.

2.4.2 Prediction sets 1 and 2: Supervised evaluation

Prediction sets 1 and 2 may be viewed as a supervised evaluation, since the TARGET sequences are labeled, i.e., they are known to be human autoepitopes, albeit not necessarily relatable to infection events. The remaining prediction sets have the entire, unlabeled human proteome (HPDS) as TARGET.

The allele set used for prediction contained all alleles for which a model exists in either CNN-PepPred or NetMHCIIpan and are present in either HADS or MEDS, totalling 58 alleles (see [Supplementary Material Table 3](#) file Alleles.xlsx).

2.4.3 Prediction sets 3 and 4: functional evaluation

Prediction sets 3 and 4 were used to investigate potential functional relationships between predicted autoantigens and the pathophysiological pathways associated with specific autoimmune diseases. This investigation involved a post-analysis of the predicted autoantigens using a systems biology approach. This approach, detailed in [Segú-Vergés et al. \(2022\)](#), evaluates the likelihood of a functional relationship between a given protein and a set of proteins based on their connections within a model of the human protein network. The underlying algorithm employs a supervised machine-learning technique, specifically Artificial Neural Networks (ANNs), trained on a large dataset comprising pharmacological drug targets and molecular descriptors of clinical phenotypes—such as drug indications and adverse effects—from the Biological Effector Database (BED ([Jorba et al., 2020](#))) compiled by the authors.

In BED, each condition (e.g., giant-cell arteritis) is characterized by a list of motifs (e.g., dysfunction of immune checkpoints), and each motif is linked to a set of proteins that map to a specific subgraph within the network. The training objective of the ANNs is to correctly associate drug targets with their respective clinical conditions. The resulting score, S , represents the probability that a given relationship is a true positive, expressed as a percentage. Here, a “relationship” refers to a scenario where a perturbation of one element (e.g., a drug target) leads to an observable perturbation in another (e.g., a clinical condition).

Although the algorithm was originally trained on drug targets rather than autoantigens, it is designed to assign probabilities to the association of any node within a protein network with a specific subgraph. Since the propagation of perturbations or signals between proteins operates on the same principles regardless of whether the context involves drug targets or other protein types, the algorithm can calculate the probability of a relationship between any protein in the network, such as a potential autoantigen, and any clinical conditions annotated in the network, such as autoimmune diseases.

Two protein-network topologies are available for the ANNs. In one of them, the functional associations are based on experimental evidence; in the other, they are based on inference and computed using a variety of resources (protein-protein interactions, gene expression, etc.). The analysis is run for both topologies and the largest score is taken. The score S for a given protein-condition/motif pair may be zero when the protein is not present in the network or no possible association is found for the pair. In such case, the pair is removed from the analysis.

In this study, S was determined for the relationship between each of the human proteins from HPDS that we predicted to be autoantigens as a consequence of microbial peptide mimicry and a list of pathophysiological motifs characteristic of each selected autoimmune disease. The list of autoimmune diseases was compiled from the same articles used for the generation of MEDS (see [Section 2.3.1](#)), and each autoimmune disease was then mapped to corresponding pathophysiological motifs compiled in BED (see [MEDS_summary.xlsx](#) in [Supplementary Material Table 1](#)).

For a given BED condition, both the separate protein sets corresponding to the individual motifs and a single protein set corresponding to all motifs of the condition were used. The motifs to be tested for each predicted autoantigen were selected using the following logical sequence: predicted autoantigen \rightarrow microbial protein with matching epitope sequence \rightarrow infectious organism \rightarrow organism's related autoimmune diseases \rightarrow BED motifs.

The distribution of S for the predicted autoantigens was compared to the distribution for a random subset of 1000 samples from HPDS, as surrogate for a random distribution, using the one-sided Mann-Whitney U test ([Di Bucchianico, 1999](#)).

2.4.4 Prediction sets 5 and 6: SARS-CoV-2 peptide mimicry

Prediction sets 5 and 6 illustrate an actual application of the tool: the identification of potential human autoantigens resulting from SARS-CoV-2 peptide mimicry.

The analysis included those HLA class II alleles for which there is experimental evidence of their binding of human autoepitopes

from HADS or SARS-CoV-2 epitopes from MEDS, plus a set of alleles from HLA-Spread (Dholakia et al., 2022) associated to autoimmune diseases that have been linked to SARS-CoV-2. As further filter, only those alleles for which there was a model in either CNN-PepPred or NetMHCIIpan were considered, leading to a total of 45 alleles (see [Supplementary Material Table 3](#) file Alleles.xlsx).

We applied the same systems biology approach used on prediction sets 3 and 4 to explore potential functional relationships between predicted autoantigens and pathophysiological pathways associated to specific autoimmune diseases. To that end, we choose the BED motifs corresponding to the following autoimmune diseases associated with SARS-CoV-2 infection (see [Supplementary Material Table 1](#) file MEDS_summary.xlsx): Anemia, Diabetes type 1, Guillain-Barre Syndrome, Myasthenia Gravis, Rheumatoid Arthritis and, Lupus Erythematosus Systemic (Ehrenfeld et al., 2020; Knight et al., 2021). For each condition motif, the background distribution of the score S was calculated using proteins that had $S > 0$, lacked a predicted autoepitope, and were not present in the motif's description. For the predicted autoantigens, the score S and its corresponding percentile, Perc(S) (indicating where it falls within the background distribution), were determined. Perc(S) served as an indicator of a potential functional relationship between the autoantigen and the pathophysiological motif. Thresholds were established as follows: Perc(S) > 95 indicated a weak functional relationship, Perc(S) > 99 indicated a functional relationship, and Perc(S) > 99.9 indicated a strong functional relationship.

3 Results and discussion

3.1 Supervised evaluation

Prediction sets 1 (using BLASTp) and 2 (using PSI-BLAST) evaluate the capacity of the tool to identify human peptides from a pool of known autoepitopes (HADS dataset) that significantly align with known epitopes from microbial species known to be associated to autoimmune diseases (MEDS dataset) and are recognised as HLA class II epitopes by one or both predictors used. In essence, starting from a pool of known human autoepitopes, the tool predicts which of them could induce autoimmunity as a consequence of a previous infection and microbial peptide mimicry. The raw results are reported in [Supplementary Material Table 4](#) file MEDS_vs_HADS.xlsx.

Out of the 807 known human autoepitops contained in HADS, 21 had at least one sequence fragment that significantly aligned with a microbial epitope from MEDS and was predicted as an autoepitope by CNN-PepPred and/or NetMHCIIpan ([Table 1](#)). The matching microbial epitopes are from SARS-CoV-2, *Mycobacterium tuberculosis* (MT) and Human Alphaherpesvirus (HHV) 1 and 3. PSI-BLAST and BLASTp produced significant alignments in all cases.

In four of the cases, the predicted autoepitopes were matched at the epitope-allele level and using OneHLA as allele selection criteria (see [Section 2.4.1](#)), meaning that the predictions matched the allele that has been found to bind, experimentally, both the microbial and human epitopes. The corresponding autoantigens are H1-2, H1-4 and MPO, showing sequence similarity with HbhA and RplV from MT and the Spike Glicoprotein from SARS-CoV-2 ([Table 1](#)).

Furthermore, the alleles DRB1*15:01 and DRB5*01:01 are known to be linked to the autoimmune disease—multiple sclerosis—that has been associated with these autoantigens (Karni et al., 1999; Finn et al., 2004; Prat et al., 2005; Zivković et al., 2009; Shahbazi et al., 2010; Alcina et al., 2012; Quandt et al., 2012; Apperson et al., 2013; Kaushansky and Ben-Nun, 2014; Stürner et al., 2019).

3.2 Functional evaluation

Prediction sets 3 (using BLASTp) and 4 (using PSI-BLAST) evaluate the capacity of the tool to identify peptides from the full human proteome (HPDS dataset) that significantly align with known epitopes from microbial species known to be associated to autoimmune diseases (MEDS dataset) and are recognised as HLA class II epitopes by one or both predictors used. In short, it extends the analysis reported in [Section 3.1](#) to the full human proteome. The raw results are reported in [Supplementary Material Table 5](#) file MEDS_vs_HPDS.xlsx.

The identified autoantigens were then subjected to the functional evaluation described in [Section 2.4.3](#). The results of this evaluation are reported in [Supplementary Material Table 6](#) file MEDS_functional.xlsx. The results are summarized in [Figure 2](#) and [Table 2](#), where the dependence of the distribution of scores S —the probability that there exists a relationship between the predicted autoantigen and a pathophysiological pathway associated to one of the autoimmune diseases considered—on the various parameters—E-value and %Rank thresholds, use of BLASTp or PSI-BLAST, use of the allHLA or oneHLA allele-selection criteria—is evaluated. In addition, the different score distributions are compared to a score distribution for a random subset of 1000 proteins from HPDS.

Using PSI-BLAST notably reduces the number of hits, but its strength lies in incorporating specific biological information into the alignments. By leveraging the pool of peptides known to bind HLA class II molecules to generate a Position-Specific Scoring Matrix, PSI-BLAST may enhance the alignment's relevance. Although this approach results in fewer hits, it achieves the highest mean score ratio and the most significant distinction between predicted autoantigen score distributions and the random peptide score distribution, particularly at E-value and %Rank thresholds of 0.01 and 2, respectively. These thresholds optimize the selection of predicted autoantigens, increasing the likelihood of identifying proteins associated with a specific autoimmune disease in the human protein network. However, using a more stringent E-value threshold of 0.001 drastically reduces the number of aligned sequences, which, when combined with a %Rank threshold of 2, can prevent the prediction of any autoantigens. The choice of allele-selection criteria for the binding prediction, allHLA or oneHLA, plays also a significant role in the results, particularly regarding the number of hits.

The results of this analysis require some biological context for proper interpretation. We are examining the potential relationship, within a graph representing the human protein network, between a protein identified as a potential autoantigen and a set of proteins that have been linked to the pathophysiology of a specific autoimmune disease. Clearly, the proteins in this set tend to be elements of the immune system. Therefore, a high score indicates that the predicted

autoantigen is, in network terms, associated to these immune-system elements. While this may not be a universal characteristic of all autoantigens, Figure 2 and Table 2 show that as we apply more stringent alignment and epitope prediction thresholds, thereby increasing our confidence in the autoepitope, the mean score and significance of the relationship between autoantigen and autoimmune disease motif also increase. This focused approach not only enhances the reliability of our predictions but also allows us to propose potential links between the predicted autoantigens and specific autoimmune diseases, which would be difficult to establish otherwise.

3.3 SARS-CoV-2 peptide mimicry

3.3.1 The predicted autoantigens and their relation to autoimmune diseases

Prediction sets 5 (using BLASTp) and 6 (using PSI-BLAST) evaluate the capacity of the tool to identify peptides from the full human proteome (HPDS dataset) that significantly align with known epitopes from SARS-CoV-2 (SC2DS dataset) and are recognised as HLA class II epitopes by one or both predictors used. It thus focuses the analysis on finding SARS-CoV-2 epitopes that could induce an autoimmune disease through peptide mimicry. The raw results are reported in Supplementary Material Table 7 file SC2DS_vs_HPDS.xlsx. As in the previous section, the identified autoantigens were evaluated for their potential relationship with autoimmune-disease motifs in the human protein network (Section 2.4.3). The results of this evaluation are reported in Supplementary Material Table 8 file SC2DS_functional.xlsx.

Using BLASTp, MMPred identified 14 potential autoantigens: MYT1L, BAZ1A, CHD5, MCM8, ATF7, MOV10, MOV10L1, DNA2, BRI3, PARVG, CALD1, MICAL3, SLC35E4, and UNC50, using a threshold E-value of 0.01. In contrast, PSI-BLAST did not detect any autoepitopes with an E-value below 0.01 and predicted only one autoepitope from HELZ2 with an E-value below 0.05. This outcome is consistent with the previous section's analysis, where BLASTp yielded a significantly higher number of positive predictions at the same thresholds. Most alignments involved SARS-CoV-2 Non-Structural Proteins (NSPs), particularly NSP3, NSP5, NSP13, NSP14, NSP15, and NSP16. Additionally, one alignment involved Nucleoprotein N, and two alignments were related to the Spike protein of the Omicron variants BA.1-like and BA.4-like (Wu et al., 2020). These findings are summarized in Table 3.

In the functional analysis, the 15 predicted autoantigens were assessed for potential associations with 32 autoimmune-disease motifs annotated in the network. Eight proteins—BAZ1A, ATF7, MOV10, DNA2, PARVG, MICAL3, SLC35E4, and HELZ2—scored in the 95th percentile (Perc(S), see Section 2.4.4) or higher for 19 different motifs (see Figure 3). While no autoantigen met the minimum threshold for certain motifs, at least one motif from each selected autoimmune disease had a significant hit. Notably, BAZ1A, MOV10, and PARVG were the only proteins with a Perc(S) exceeding 99 for at least one motif. The highest Perc(S) (above 99.9) was achieved by MOV10 in association with the “Lupus Erythematosus Systemic” motif.

With a Perc(S) > 95, Rheumatoid Arthritis and Thyroiditis are associated with the same predicted autoantigens, MOV10 and ATF7, which are also linked to Lupus Erythematosus Systemic along with PARVG. Additionally, SLC35E4 is also connected to Rheumatoid Arthritis. Type I Diabetes is associated with the predicted autoantigens MOV10 and PARVG, with PARVG also linked to Guillain-Barré Syndrome, alongside ATF7, BAZ1A, and DNA2. Myasthenia Gravis shows predicted associations with HELZ2, ATF7 and PARVG. Lastly, Anemia is associated with MICAL3 and BAZ1A.

When the threshold is raised to Perc(S) > 99, MOV10 is associated with both Lupus Erythematosus Systemic and Rheumatoid Arthritis. MOV10 is also linked to Thyroiditis, along with ATF7. Anemia is connected to BAZ1A, and Myasthenia Gravis is linked to PARVG.

The human proteins MYT1L, CHD5, MCM8, MOV10L1, BRI3 and CALD1 were predicted as autoantigens but did not show a Perc(S) > 95 for any of the motifs tested.

3.3.2 The predicted human autoepitopes align with known SARS-CoV-2 epitopes

While there is no experimental evidence in IEDB linking our predicted autoepitopes to autoimmune diseases, many of the SARS-CoV-2 sequences that align significantly with these predicted autoepitopes are known to bind HLA class II molecules. Specifically, the SARS-CoV-2 sequences 369-378 of the Nucleoprotein, 47-55, 275-289, and 557-569 of NSP13, 389-398 of NSP14, 248-258 of NSP15, and 257-267 of NSP16 overlap with regions that have been experimentally validated (Obermaier et al., 2022; Huisman et al., 2022) (see Table 3). Notably, the regions 275-289 and 557-569 of NSP13 show allele-specific hits, where the autoepitopes have been predicted for the same alleles experimentally observed to bind these SARS-CoV-2 protein regions. Thus, DRB1*07:01 is predicted to bind an autoepitope in MOV10, and DRB1*04:04 is predicted to bind an autoepitope in HELZ2.

The functional analysis reveals an interesting pattern among the predicted autoantigens that align with known SARS-CoV-2 epitopes. Specifically, nine human proteins—ATF7, HELZ2, MOV10, MOV10L1, DNA2, BRI3, PARVG, CALD1, MICAL3—exhibit significant alignment with these SARS-CoV-2 epitopes. Among these, ATF7, HELZ2, MOV10, DNA2, PARVG and MICAL3 rank in the 95th percentile or higher (Perc(S) > 95) across 18 different motifs, with ATF7, MOV10, and PARVG showing Perc(S) > 99 for six motifs. MOV10 alone achieves a Perc(S) > 99.9 for a single motif. In contrast, the associations are less significant when examining predicted autoantigens derived from alignments with SARS-CoV-2 sequences not known to be antigenic. Of the six proteins with peptides matching these regions (MYT1L, BAZ1A, CHD5, MCM8, SLC35E4 and UNC50), only BAZ1A and SLC35E4 yield significant results, both achieving Perc(S) > 95 for just three motifs, with only BAZ1A reaching Perc(S) > 99 for a single motif.

3.3.3 The antiviral activity of the predicted autoantigens MOV10 and HELZ2

We examined baseline expression patterns of the predicted autoantigens in lung cells using the Expression Atlas (George et al., 2024). Proteomics data from Wang et al. (2019) indicate

that 11 out of the 15 predicted autoantigens (BAZ1A, MCM8, ATF7, HELZ2, MOV10, MOV10L1, DNA2, PARVG, CALD1, MICAL3, UNC50) are highly expressed, supporting the hypothesis that cross-reaction with HLA class II molecules in previously infected lung cells is plausible (Kawasaki et al., 2022; Hoffmann et al., 2020). Additionally, a gene enrichment analysis conducted using the online tool g:Profiler (Kolberg et al., 2023) (threshold = 0.05, multiple hypothesis testing method g:SCS) revealed significant enrichment for the Gene Ontology (GO, <https://geneontology.org/>) Molecular Function terms “helicase activity” (GO:0004386) and “single-stranded DNA-helicase” (GO:0017116), as well as the Cellular Component terms “P granule” (GO:0043186) and “intracellular non-membrane-bounded organelle” (GO:0043232). Notably, the human proteins MOV10, HELZ2, and DNA2, which are reported to have helicase activity with GO evidence code “inferred by direct assay” (IDA), contain segments that align with the helicase NSP13 of SARS-CoV-2 (see Table 3). MOV10 is implicated in the modulation of viral infectivity (Goodier et al., 2012) and promotes type I interferon production (Cuevas et al., 2016; Yang et al., 2022; Balinsky et al., 2017), while HELZ2 is known to respond to interferon production during viral infection (Huntzinger et al., 2023; Du et al., 2024). DNA2 does not appear to have any known antiviral activity. On the other hand, the helicase activity of NSP13 is crucial for viral replication (Yan et al., 2021) and this protein interacts with the host to inhibit interferon-beta production, thereby evading the immune response (Xia et al., 2020). Thus, it seems that MOV10 and HELZ2 play roles antagonistic to that of NSP13.

Overall, transcriptomic and proteomic data on our predicted autoantigens in the context of SARS-CoV-2 infection primarily focus on MOV10 and HELZ2. These studies utilize samples obtained from either the lungs of infected patients or lung cell lines.

Wang et al. employed Dermatan Sulfate (DS)-affinity proteomics to define the autoantigenome of lung fibroblasts, complemented by bioinformatics analyses to explore the relationship between autoantigenic proteins and COVID-19-induced alterations (Wang et al., 2022). Notably, they discovered that 86% of their predicted autoantigens were either up- or downregulated in COVID-19 patients or SARS-CoV-2-infected cells. Among the previously unknown autoantigens identified in this study, MOV10 (with very high DS affinity) and CALD1 (with medium to high DS affinity) align with our predictions. Both proteins exhibited altered expression in COVID-19 patients and/or SARS-CoV-2-infected cells.

In a study by An et al., the authors developed a bioinformatics pipeline similar to ours (as noted in the Introduction) and conducted a differential expression analysis using the Calu-3 human lung adenocarcinoma cell line (An et al., 2022). Their pipeline predicts that MOV10 and HELZ2 may exhibit cross-reactivity, while expression data reveal a positive correlation between transcript abundance of these proteins, particularly HELZ2, and SARS-CoV-2 viral load in infected cells.

A study by Ariumi explored the epigenetic mechanisms triggered by SARS-CoV-2 infections (Ariumi, 2022). The findings indicate that the knockdown of MOV10 leads to a significant increase in viral load/replication in infected cells, suggesting that MOV10 plays a role in the host's suppression of SARS-CoV-2 replication.

A study examining gene expression in cell lines and patient samples in the context of epigenetic regulation during SARS-CoV-2, SARS-CoV, and MERS infections identified various differentially expressed genes involved in the epigenetic response during infection in pulmonary cell lines (Salgado-Albarrán et al., 2021). Although the study does not report expression data for MOV10, it highlights MOV10's functional and physical relationships with the differentially expressed genes through protein-protein interaction (PPI) data (Kotlyar et al., 2019) and co-expression analysis (Langfelder and Horvath, 2008).

Aside from MOV10 and HELZ2, the literature also reports the overexpression of MICAL3 in convalescent COVID-19 patients who retested positive (Fang et al., 2022). However, no relevant studies have been found that link the other predicted autoantigens to SARS-CoV-2.

3.3.4 Sequence and structural similarity between NSP13 and MOV10, an immune-escape mechanism?

Considering the reported evidence regarding MOV10, we extended our investigation to NSP13 due to its shared helicase activity and the presence of a similar HLA class II epitope. Given MOV10's suggested suppressor role in SARS-CoV-2 replication, its presence in lung epithelial cells, its upregulated expression in SARS-CoV-2-infected cells, and the APC-like properties of lung epithelial cells (Kawasaki et al., 2022; Hoffmann et al., 2020), we propose that cross-reactive epitopes in NSP13 could potentially trigger autoimmunity and facilitate SARS-CoV-2 replication in the lungs.

We would therefore expect the NSP13 epitope NVNRFNVAITRAK (positions 557 to 569, see Table 3) to be conserved across NSP13 variants. However, it is important to note that this conservation might also arise from the sequence's involvement in NSP13's catalytic activity (Newman et al., 2021).

A total of 1,725,419 protein sequences of the ORF1ab polyproteins were downloaded from NCBI Virus (15/02/2024) (Hatcher et al., 2017), with the maximum number of ambiguous characters set to zero, and considering only sequences collected for baseline surveillance (random sampling = “Only”). To extract NSP13 from the ORF1ab sequences, a BLASTp alignment was performed using the reference sequence of NSP13 as the query and the ORF1ab sequences as the target. Due to the high similarity of the variants to the reference, all alignments yielded significant results, allowing us to heuristically extract the corresponding NSP13 variants. A multiple sequence alignment of all NSP13 variants, along with the NSP13 reference, was then performed using FAMSA (Deorowicz et al., 2016), which is optimized for datasets with high dimensionality and high pairwise identity. To compute the conservation of the NSP13 epitope, we averaged the Shannon Entropy (Shenkin et al., 1991) across its positions and compared it to the distribution of all windows of the same length across the alignment. The NSP13 epitope at positions 557 to 569 showed a higher conservation score than 95.7% of the windows of the same size, indicating the strong conservation of this region. A logo plot of the alignment is presented in Figure 4.

Furthermore, the superposition of the crystallographic structure of NSP13 (PDB entry 6ZSL, chain B) (Newman et al., 2021) with the predicted AlphaFold structure of MOV10 (UniProt entry Q9HCE1) (Jumper et al., 2021; Varadi et al., 2022) reveals significant

structural similarity between two regions of these proteins of 214 residues in length (36.5% of the sequence of 6ZSL and 21.3% of that of the AlphaFold model of Q9HCE1, with a C_{α} RMSD of 1.98 Å, see Figure 4). Notably, the regions corresponding to the epitopes—NSP13 residues 557–569 and MOV10 residues 901–913—are superimposed in the structural alignment, suggesting that epitope mimicry could arise from both sequence and structure. Although the MOV10 structure is based on an AlphaFold model (Terwilliger et al., 2024), the model confidence in this region is very high (pLDDT > 90).

The molecular mimicry mechanism hypothetically employed by SARS-CoV-2 may parallel other viral strategies identified as triggers for autoimmune diseases (Zhao et al., 1998; Hrycek et al., 2005; Bjernevik et al., 2022), functioning in two potential ways as outlined by Maguire et al. (2024): *i*) by mimicking a structural region of MOV10, NSP13 could enable SARS-CoV-2 to evade immune detection and achieve immune tolerance, facilitating viral persistence and diminishing activation of the host's adaptive immune response (Maguire et al., 2024; Cusick et al., 2011; Zhao et al., 1998; Adiguzel, 2021); *ii*) NSP13's sequence and structural resemblance to MOV10 might competitively inhibit MOV10's activity, impairing its antiviral function in lung epithelial cells and promoting SARS-CoV-2 replication. Additionally, the similarity between NSP13 and MOV10 may induce the development of cross-reactive T-cells and/or antibodies via HLA class II epitope recognition, potentially triggering an autoimmune response (Cusick et al., 2011; Smatti et al., 2019; Zhao et al., 1998; Sabbatini et al., 1993).

It is noteworthy that these mechanisms could operate concurrently, with NSP13's epitope conservation enhancing SARS-CoV-2 infection in both scenarios. The predominance of one mechanism over the other likely depends on the specific HLA class II alleles present, as different alleles are associated with distinct outcomes (Dholakia et al., 2022).

Overall, these findings support the hypothesis of cross-reactivity between the epitopes of NSP13 and MOV10, which may influence SARS-CoV-2 replication in the lungs. While the potential for an autoimmune response against MOV10 and its connection to autoimmune diseases remains speculative at this stage (see Section 3.3.1), further investigation is warranted.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2016). TensorFlow: large-scale machine learning on heterogeneous distributed systems. *arXiv*, 1603. doi:10.48550/arXiv.1603.04467
- Adiguzel, Y. (2021). Molecular mimicry between SARS-CoV-2 and human proteins. *Autoimmune. Rev.* 20, 102791. doi:10.1016/j.autrev.2021.102791
- Alcina, A., Abad-Grau, M. M., Fedetz, M., Izquierdo, G., Lucas, M., Fernández, O., et al. (2012). Multiple sclerosis risk variant HLA-DRB1*1501 associates with high expression of DRB1 gene in different human populations. *PLoS One* 7, e29819. doi:10.1371/journal.pone.0029819
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2
- An, H., Eun, M., Yi, J., and Park, J. (2022). CRESSP: a comprehensive pipeline for prediction of immunopathogenic SARS-CoV-2 epitopes using structural properties of proteins. *Brief. Bioinform.* 23, bbac056. doi:10.1093/bib/bbac056
- Apperson, M. L., Tian, Y., Stamova, B., Ander, B. P., Jickling, G. C., Agius, M. A., et al. (2013). Genome wide differences of gene expression associated with HLA-DRB1 genotype in multiple sclerosis: a pilot study. *J. Neuroimmunol.* 257, 90–96. doi:10.1016/j.jneuroim.2013.02.004

Author contributions

FG: Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing—original draft. VJ: Investigation, Software, Writing—review and editing. JF: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing—review and editing. XD: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing—original draft, Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreements No. 765158 (COSMIC) and 813545 (HELICAL) and from the Spanish Ministry for Science, Innovation and Universities under grant PID2019-111364RB-I00.

Conflict of interest

Authors FG, VJ and JF are employed by Anaxomics Biotech, Barcelona, Spain.

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2024.1500684/full#supplementary-material>

- Ariumi, Y. (2022). Host cellular RNA helicases regulate SARS-CoV-2 infection. *J. Virol.* 96, e0000222. doi:10.1128/jvi.00002-22
- Balinsky, C. A., Schmeisser, H., Wells, A. I., Ganesan, S., Jin, T., Singh, K., et al. (2017). IRAG (FLJ11286), an interferon-stimulated gene with antiviral activity against dengue virus, interacts with MOV10. *J. Virol.* 91, e01606-16–e01616. doi:10.1128/JVI.01606-16
- Bjornevik, K., Cortese, M., Healy, B. C., Kuhle, J., Mina, M. J., Leng, Y., et al. (2022). Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science* 375, 296–301. doi:10.1126/science.abj8222
- Caron, E., Kowalewski, D. J., Chiek Koh, C., Sturm, T., Schuster, H., and Aebersold, R. (2015). Analysis of Major Histocompatibility Complex (MHC) immunopeptidomes using mass spectrometry. *Mol. Cell. Proteomics* 14, 3105–3117. doi:10.1074/mcp.O115.052431
- Cuevas, R. A., Ghosh, A., Wallerath, C., Hornung, V., Coyne, C. B., and Sarkar, S. N. (2016). MOV10 provides antiviral activity against RNA viruses by enhancing RIG-I - MAVS-independent IFN induction. *J. Immunol.* 196, 3877–3886. doi:10.4049/jimmunol.1501359
- Cunningham, M. W. (2014). Rheumatic fever, autoimmunity and molecular mimicry: the streptococcal connection. *Int. Rev. Immunol.* 33, 314–329. doi:10.3109/08830185.2014.917411
- Cusick, M. F., Libbey, J. E., and Fujinami, R. S. (2011). Molecular mimicry as a mechanism of autoimmune disease. *Clin. Rev. Allergy Immunol.* 42, 102–111. doi:10.1007/s12016-011-8294-7
- Deorowicz, S., Debudaj-Grabysz, A., and Gudyś, A. (2016). FAMSA: fast and accurate multiple sequence alignment of huge protein families. *Sci. Rep.* 6, 33964. doi:10.1038/srep33964
- Dholakia, D., Kalra, A., Misir, B. R., Kanga, U., and Mukerji, M. (2022). HLA-SPREAD: a natural language processing based resource for curating HLA association from PubMed abstracts. *BMC Genomics* 23, 10. doi:10.1186/s12864-021-08239-0
- Di Bucchianico, A. (1999). Combinatorics, computer algebra and the Wilcoxon-Mann-Whitney test. *J. Stat. Plan. Inference* 79, 349–364. doi:10.1016/S0378-3758(98)00261-4
- Doxey, A. C., and McConkey, B. J. (2013). Prediction of molecular mimicry candidates in human pathogenic bacteria. *Virulence* 4, 453–466. doi:10.4161/viru.25180
- Du, L., Deiter, F., Bouzidi, M. S., Billaud, J. N., Simmons, G., Dabral, P., et al. (2024). A viral assembly inhibitor blocks SARS-CoV-2 replication in airway epithelial cells. *Commun. Biol.* 7, 486. doi:10.1038/s42003-024-06130-8
- Ehrenfeld, M., Tincani, A., Andreoli, L., Cattalini, M., Greenbaum, A., Kanduc, D., et al. (2020). Covid-19 and autoimmunity. *Autoimmun. Rev.* 19, 102597. doi:10.1016/j.autrev.2020.102597
- Fang, K. Y., Liang, G. N., Zhuang, Z. Q., Fang, Y. X., Dong, Y. Q., Liang, C. J., et al. (2022). Screening the hub genes and analyzing the mechanisms in discharged COVID-19 patients retesting positive through bioinformatics analysis. *J. Clin. Lab. Anal.* 36, e24495. doi:10.1002/jcla.24495
- Finn, T. P., Jones, R. E., Rich, C., Dahan, R., Link, J., David, C. S., et al. (2004). HLA-DRB1*1501 risk association in multiple sclerosis may not be related to presentation of myelin epitopes. *J. Neurosci. Res.* 78, 100–114. doi:10.1002/jnr.20227
- Fiorillo, M. T., Paladini, F., Tedeschi, V., and Sorrentino, R. (2017). HLA Class I or Class II and disease association: catch the difference if you can. *Front. Immunol.* 8, 1475. doi:10.3389/fimmu.2017.01475
- George, N., Fexova, S., Fuentes, A. M., Madrigal, P., Bi, Y., Iqbal, H., et al. (2024). Expression Atlas update: insights from sequencing data at both bulk and single cell level. *Nucleic Acids Res.* 52, D107–D114. doi:10.1093/nar/gkad1021
- Goodier, J. L., Cheung, L. E., and Kazanian, H. H. (2012). MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. *PLoS Genet.* 8, e1002941. doi:10.1371/journal.pgen.1002941
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. doi:10.1038/s41586-020-2649-2
- Hatcher, E. L., Zhdanov, S. A., Bao, Y., Blinkova, O., Nawrocki, E. P., Ostapchuck, Y., et al. (2017). Virus Variation Resource - improved response to emergent viral outbreaks. *Nucleic Acids Res.* 45, D482–D490. doi:10.1093/nar/gkw1065
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280.e8. doi:10.1016/j.cell.2020.02.052
- Hrycek, A., Kuśmierz, D., Mazurek, U., and Wilczok, T. (2005). Human cytomegalovirus in patients with systemic lupus erythematosus. *Autoimmunity* 38, 487–491. doi:10.1080/08916930500285667
- Huisman, B. D., Dai, Z., Gifford, D. K., and Birnbaum, M. E. (2022). A high-throughput yeast display approach to profile pathogen proteomes for MHC-II binding. *eLife* 11, e78589. doi:10.7554/eLife.78589
- Hunter, J. D. (2007). Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95. doi:10.1109/MCSE.2007.55
- Huntzinger, E., Sinteff, J., Morlet, B., and Séraphin, B. (2023). HELZ2: a new, interferon-regulated, human 3'-5' exoribonuclease of the RNB family is expressed from a non-canonical initiation codon. *Nucleic Acids Res.* 51, 9279–9293. doi:10.1093/nar/gkad673
- Jorba, G., Aguirre-Plans, J., Junet, V., Segú-Vergés, C., Ruiz, J. L., Pujol, A., et al. (2020). In-silico simulated prototype-patients using TPMS technology to study a potential adverse effect of sacubitril and valsartan. *PLoS One* 15, e0228926. doi:10.1371/journal.pone.0228926
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2
- Junet, V., and Daura, X. (2021). CNN-PepPred: an open-source tool to create convolutional NN models for the discovery of patterns in peptide sets - application to peptide-MHC class II binding prediction. *Bioinformatics* 37, 4567–4568. doi:10.1093/bioinformatics/ctab687
- Karni, A., Kohn, Y., Safirman, C., Abramsky, O., Barcellos, L., Oksenberg, J. R., et al. (1999). Evidence for the genetic role of human leukocyte antigens in low frequency DRB1*1501 multiple sclerosis patients in Israel. *Mult. Scler. J.* 5, 410–415. doi:10.1177/1352458599005001607
- Kaushansky, N., and Ben-Nun, A. (2014). DQB1*06:02-associated pathogenic anti-myelin autoimmunity in multiple sclerosis-like disease: potential function of DQB1*06:02 as a disease-predisposing allele. *Front. Oncol.* 4, 280. doi:10.3389/fonc.2014.00280
- Kawasaki, T., Ikegawa, M., and Kawai, T. (2022). Antigen presentation in the lung. *Front. Immunol.* 13, 860915. doi:10.3389/fimmu.2022.860915
- Knight, J. S., Caricchio, R., Casanova, J. L., Combes, A. J., Diamond, B., Fox, S. E., et al. (2021). The intersection of COVID-19 and autoimmunity. *J. Clin. Invest.* 131, e154886. doi:10.1172/JCI154886
- Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., and Peterson, H. (2023). g-Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* 51, W207–W212. doi:10.1093/nar/gkad347
- Kotlyar, M., Pastrello, C., Malik, Z., and Jurisica, I. (2019). IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. *Nucleic Acids Res.* 47, D581–D589. doi:10.1093/nar/gky1037
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* 9, 559. doi:10.1186/1471-2105-9-559
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi:10.1093/bioinformatics/btl158
- Maguire, C., Wang, C., Ramasamy, A., Fonken, C., Morse, B., Lopez, N., et al. (2024). Molecular mimicry as a mechanism of viral immune evasion and autoimmunity. *Nat. Commun.* 15, 9403. doi:10.1038/s41467-024-53658-8
- Newman, J. A., Douangamath, A., Yadzani, S., Yosaatmadja, Y., Aimon, A., Brandão Neto, J., et al. (2021). Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase. *Nat. Commun.* 12, 4848. doi:10.1038/s41467-021-25166-6
- Obermair, F. J., Renoux, F., Heer, S., Lee, C. H., Cereghetti, N., Loi, M., et al. (2022). High-resolution profiling of MHC II peptide presentation capacity reveals SARS-CoV-2 CD4 T cell targets and mechanisms of immune escape. *Sci. Adv.* 8, eabl5394. doi:10.1126/sciadv.abl5394
- O'Donnell, T. J., Rubinsteyn, A., and Laserson, U. (2020). MHCflurry 2.0: improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing. *Cell Syst.* 11, 42–48.e7. doi:10.1016/j.cels.2020.06.010
- Oldstone, M. B. A. (1998). Molecular mimicry and immune-mediated diseases. *FASEB J.* 12, 1255–1265. doi:10.1096/fasebj.12.13.1255
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. doi:10.5555/1953048.2078195
- Prat, E., Tomaru, U., Sabater, L., Park, D. M., Granger, R., Kruse, N., et al. (2005). HLA-DRB5*0101 and -DRB1*1501 expression in the multiple sclerosis-associated HLA-DR15 haplotype. *J. Neuroimmunol.* 167, 108–119. doi:10.1016/j.jneuroim.2005.04.027
- Quandt, J. A., Huh, J., Baig, M., Yao, K., Ito, N., Bryant, M., et al. (2012). Myelin basic protein-specific TCR/HLA-DRB5*01:01 transgenic mice support the etiologic role of DRB5*01:01 in multiple sclerosis. *J. Immunol.* 189, 2897–2908. doi:10.4049/jimmunol.1103087
- Rambaut, A., Holmes, E. C., O'Toole, A., Hill, V., McCrone, J. T., Ruis, C., et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 5, 1403–1407. doi:10.1038/s41564-020-0770-5
- Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* 48, W449–W454. doi:10.1093/nar/gkaa379
- Rojas, M., Restrepo-Jiménez, P., Monsalve, D. M., Pacheco, Y., Acosta-Ampudia, Y., Ramírez-Santana, C., et al. (2018). Molecular mimicry and autoimmunity. *J. Autoimmun.* 95, 100–123. doi:10.1016/j.jaut.2018.10.012

- Sabbatini, A., Bombardieri, S., and Migliorini, P. (1993). Autoantibodies from patients with systemic lupus erythematosus bind a shared sequence of SmD and Epstein-Barr virus-encoded nuclear antigen EBNA 1. *Eur. J. Immunol.* 23, 1146–1152. doi:10.1002/eji.1830230525
- Salgado-Albarrán, M., Navarro-Delgado, E. I., Del Moral-Morales, A., Alcaraz, N., Baumbach, J., González-Barrios, R., et al. (2021). Comparative transcriptome analysis reveals key epigenetic targets in SARS-CoV-2 infection. *npj Syst. Biol. Appl.* 7, 21. doi:10.1038/s41540-021-00181-x
- Schäffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., et al. (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* 29, 2994–3005. doi:10.1093/nar/29.14.2994
- Segú-Vergés, C., Caño, S., Calderón-Gómez, E., Bartra, H., Sardón, T., Kaveri, S., et al. (2022). Systems biology and artificial intelligence analysis highlights the pleiotropic effect of IVIg therapy in autoimmune diseases with a predominant role on B cells and complement system. *Front. Immunol.* 13, 901872. doi:10.3389/fimmu.2022.901872
- Shahbazi, M., Roshandel, D., Ebadi, H., Fathi, D., Zamani, M., Boghaee, M., et al. (2010). High frequency of the IL-2 - 330T/HLA-DRB1*1501 haplotype in patients with multiple sclerosis. *Clin. Immunol.* 137, 134–138. doi:10.1016/j.clim.2010.05.010
- Shenkin, P. S., Erman, B., and Mastrandrea, L. D. (1991). Information-theoretical entropy as a measure of sequence variability. *Proteins* 11, 297–313. doi:10.1002/prot.340110408
- Smatti, M. K., Cyprian, F. S., Nasrallah, G. K., Al Thani, A. A., Almishal, R. O., and Yassine, H. M. (2019). Viruses and autoimmunity: a review on the potential interaction and molecular mechanisms. *Viruses* 11, 762. doi:10.3390/v11080762
- Stürner, K. H., Siembab, I., Schön, G., Stellmann, J. P., Heidari, N., Fehse, B., et al. (2019). Is multiple sclerosis progression associated with the HLA-DR15 haplotype? *Mult. Scler. J. Exp. Transl. Clin.* 5, 2055217319894615. doi:10.1177/2055217319894615
- Tareen, A., and Kinney, J. B. (2020). Logomaker: beautiful sequence logos in Python. *Bioinformatics* 36, 2272–2274. doi:10.1093/bioinformatics/btz921
- Terwilliger, T. C., Liebschner, D., Croll, T. I., Williams, C. J., McCoy, A. J., Poon, B. K., et al. (2024). AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nat. Methods* 21, 110–116. doi:10.1038/s41592-023-02087-4
- The UniProt Consortium (2023). UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* 51, D523–D531. doi:10.1093/nar/gkac1052
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439–D444. doi:10.1093/nar/gkab1061
- Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., et al. (2019). The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res.* 47, D339–D343. doi:10.1093/nar/gky1006
- Wang, D., Eraslan, B., Wieland, T., m, B., Hopf, T., Zolg, D. P., et al. (2019). A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* 15, e8503. doi:10.15252/msb.20188503
- Wang, J. Y., Zhang, W., Roehrl, V. B., Roehrl, M. W., and Roehrl, M. H. (2022). An autoantigen atlas from human lung HFL1 cells offers clues to neurological and diverse autoimmune manifestations of COVID-19. *Front. Immunol.* 13, 831849. doi:10.3389/fimmu.2022.831849
- Wang, P., Sidney, J., Kim, Y., Sette, A., Lund, O., Nielsen, M., et al. (2010). Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinforma.* 11, 568. doi:10.1186/1471-2105-11-568
- Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi:10.1038/s41586-020-2008-3
- Wu, Y., Zhang, N., Hashimoto, K., Xia, C., and Dijkstra, J. M. (2021). Structural comparison between MHC classes I and II; in evolution, a class-II-like molecule probably came first. *Front. Immunol.* 12, 621153. doi:10.3389/fimmu.2021.621153
- Xia, H., Cao, Z., Xie, X., Zhang, X., Chen, J. Y., Wang, H., et al. (2020). Evasion of type I interferon by SARS-CoV-2. *Cell Rep.* 33, 108234. doi:10.1016/j.celrep.2020.108234
- Yan, L., Ge, J., Zheng, L., Zhang, Y., Gao, Y., Wang, T., et al. (2021). Cryo-EM structure of an extended SARS-CoV-2 replication and transcription complex reveals an intermediate state in cap synthesis. *Cell* 184, 184–193.e10. doi:10.1016/j.cell.2020.11.016
- Yang, X., Xiang, Z., Sun, Z., Ji, F., Ren, K., and Pan, D. (2022). Host MOV10 is induced to restrict herpes simplex virus 1 lytic infection by promoting type I interferon response. *PLoS Pathog.* 18, e1010301. doi:10.1371/journal.ppat.1010301
- Zhao, Z. S., Granucci, F., Yeh, L., Schaffer, P. A., and Cantor, H. (1998). Molecular mimicry by herpes simplex virus-type 1: autoimmune disease after viral infection. *Science* 279, 1344–1347. doi:10.1126/science.279.5355.1344
- Zivković, M., Stanković, A., Dinčić, E., Popović, M., Popović, S., Raičević, R., et al. (2009). The tag SNP for HLA-DRB1*1501, rs3135388, is significantly associated with multiple sclerosis susceptibility: cost-effective high-throughput detection by real-time PCR. *Clin. Chim. Acta* 406, 27–30. doi:10.1016/j.cca.2009.05.004