



OPEN ACCESS

EDITED BY

Helder Nakaya,
University of São Paulo, Brazil

REVIEWED BY

Ravi Kumar,
Central University of Haryana, India
Sudeep Kumar Maurya,
University of Pittsburgh Medical Center,
United States

*CORRESPONDENCE

Li C. Xue

✉ me.lixue@gmail.com

Mohammad Hossein Karimi-Jafari

✉ mhkarimijafari@ut.ac.ir

†These authors have contributed equally to this work

RECEIVED 30 August 2023

ACCEPTED 17 November 2023

PUBLISHED 08 December 2023

CITATION

Parizi FM, Marzella DF, Ramakrishnan G, 't Hoen PAC, Karimi-Jafari MH and Xue LC (2023) PANDORA v2.0: Benchmarking peptide-MHC II models and software improvements. *Front. Immunol.* 14:1285899. doi: 10.3389/fimmu.2023.1285899

COPYRIGHT

© 2023 Parizi, Marzella, Ramakrishnan, 't Hoen, Karimi-Jafari and Xue. 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.

PANDORA v2.0: Benchmarking peptide-MHC II models and software improvements

Farzaneh M. Parizi^{1,2†}, Dario F. Marzella^{1†},
Gayatri Ramakrishnan¹, Peter A. C. 't Hoen¹,
Mohammad Hossein Karimi-Jafari^{2*} and Li C. Xue^{1*}

¹Medical BioSciences Department, Radboud University Medical Center, Nijmegen, Netherlands,

²Department of Bioinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

T-cell specificity to differentiate between self and non-self relies on T-cell receptor (TCR) recognition of peptides presented by the Major Histocompatibility Complex (MHC). Investigations into the three-dimensional (3D) structures of peptide:MHC (pMHC) complexes have provided valuable insights of MHC functions. Given the limited availability of experimental pMHC structures and considerable diversity of peptides and MHC alleles, it calls for the development of efficient and reliable computational approaches for modeling pMHC structures. Here we present an update of PANDORA and the systematic evaluation of its performance in modelling 3D structures of pMHC class II complexes (pMHC-II), which play a key role in the cancer immune response. PANDORA is a modelling software that can build low-energy models in a few minutes by restraining peptide residues inside the MHC-II binding groove. We benchmarked PANDORA on 136 experimentally determined pMHC-II structures covering 44 unique $\alpha\beta$ chain pairs. Our pipeline achieves a median backbone Ligand-Root Mean Squared Deviation (L-RMSD) of 0.42 Å on the binding core and 0.88 Å on the whole peptide for the benchmark dataset. We incorporated software improvements to make PANDORA a pan-allele framework and improved the user interface and software quality. Its computational efficiency allows enriching the wealth of pMHC binding affinity and mass spectrometry data with 3D models. These models can be used as a starting point for molecular dynamics simulations or structure-boosted deep learning algorithms to identify MHC-binding peptides. PANDORA is available as a Python package through Conda or as a source installation at <https://github.com/X-lab-3D/PANDORA>.

KEYWORDS

peptide:MHC, MHC class II, peptide binding, 3D structures, large-scale 3D modelling

1 Introduction

The ability of T-cells to recognize and eliminate infected or transformed cells relies on their ability to distinguish between self and non-self peptides presented by the Major Histocompatibility Complex (MHC) on the surface of these cells. Upon recognition of a non-self peptide by T-cell receptors (TCR), T-cells activate and initiate an immune response. MHC class I (MHC-I) molecules typically present intracellular antigens to cytotoxic CD8+ T-cells, which eliminate the cell presenting the antigen. MHC-II molecules present extracellular antigens to helper CD4+ T-cells, which assist other immune cells by releasing cytokines and orchestrating the immune response (1, 2). To unravel the mechanisms of peptide presentation to T-cells and immune response, it is essential to investigate how peptides bind to MHC molecules.

Understanding the mechanism of peptide-MHC (pMHC) binding raises an intriguing research question regarding how MHC molecules effectively bind to a wide range of peptides while maintaining strong binding and specificity. Previous research focusing on the structural aspects of pMHC complexes has provided valuable insights into our understanding of antigen presentation specificity (3) and peptide binding dynamics (4, 5). Allele-specific residues at anchor positions and complementary pockets in the MHC molecule play a significant role in determining the promiscuity and specificity of peptide recognition by MHC molecules (6, 7). Notably, the presence of hydrophobic anchors and the formation of hydrogen bonds have been discovered to stabilize the pMHC-II interaction (8, 9). Similarly, in the case of MHC class I, peptide-dependent stability is achieved through the establishment of conserved hydrogen bonds at the N and C termini of peptides, along with anchor residues that fit into pockets of MHC class I (10, 11). Furthermore, structural investigations have provided insights into other mechanisms, such as the molecular basis of autoimmune diseases (10) and T-cell recognition (11, 12). The knowledge gained from structural studies has also facilitated the design of novel therapies and can help the development of

effective vaccine strategies (13, 14). Therefore, access to structural information on pMHC is crucial for these advancements.

This work focuses on pMHC-II binding. MHC-II is crucial in antigen presentation, particularly for extracellular antigens. Additionally, MHC-II mediated CD4+ T-cell responses are reported to account for the predominant immune responses following cancer vaccine treatment (15–18). The MHC-II complex consists of two membrane-anchored chains: an α - and β -chain (Figure 1), and it can bind peptides up to 25 residues in length (20, 21). The binding groove of MHC-II can hold a 9-mer core (22). The residues outside the groove form the Peptide Flanking Regions (PFR), namely the left (N-terminus) and right (C-terminus) PFRs. A peptide is kept in place within the groove by three or four main conserved binding pockets: Pockets 1, 4, 6, and 9 (Figure 1A, and alongside these, there are smaller auxiliary anchor pockets (23, 24).

To accommodate a diverse range of antigens within the MHC groove, the MHC locus stands out as the most polymorphic region in the human genome (2, 25). With over 10,754 alleles for MHC class II, there is a significant variation in MHC-II alleles and the peptides they can bind (26). Unfortunately, only a few pMHC-II structures have been experimentally resolved [about 240 entries in the PDB, the Protein Data Bank (27)]. This necessitates the development of fast, structure-based computational modeling methods to overcome the scarcity of available pMHC-II structures. However, only a few modeling methods have been explicitly developed for pMHC-II complexes.

Most existing pMHC-II modelling methods rely on grid-based docking, including pDock and EpiDock (28–33). Among them, pDock has demonstrated improved performance in generating peptide core conformations bound to MHC-II. The pDock's approach involves receptor modeling followed by flexible peptide docking into the binding groove while retaining its starting conformation using loose restraints. Current pMHC-II modeling approaches are often limited in terms of usability due to: 1) long computation times; 2) the use of closed-source software; 3) limited

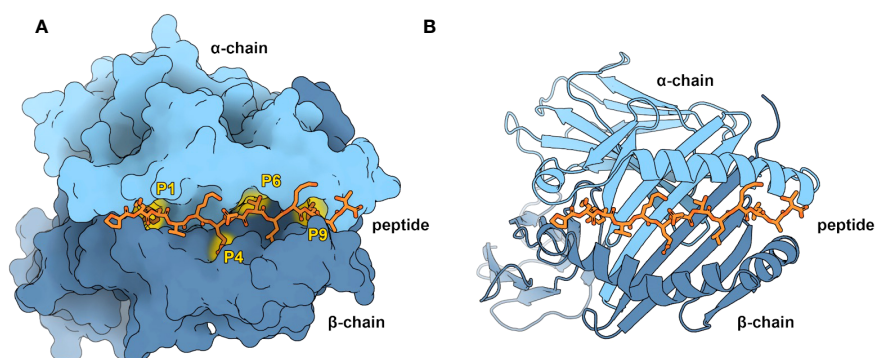


FIGURE 1

Overview of the pMHC-II complex (A) Representation of an MHC-II molecule by its accessible surface area, visualized with Protein Imager (19). MHC-II consists of an α -chain (light blue) and a β -chain (dark blue). Shown are the four characteristic pockets in the binding groove (P1, P4, P6, and P9), occupied by the corresponding peptide (orange) anchor residues. As modeling restraints (yellow), PANDORA uses the atomic contacts between the peptide anchor residues and the MHC-II pockets. (B) Cartoon representation of pMHC-II. The peptide binding groove consists of two α -helices on a floor of β -sheets, in which the peptide resides (PDB ID: 1DLH).

coverage of diverse MHC alleles; and 4) uncertainty regarding the quality of PFR conformations. Additionally, structural modelling of pMHC-II complexes is fraught with challenges. It is not always clear which region of a peptide forms the core and is directly anchored to the MHC-II receptors (34, 35). Existing methods, such as NetMHCIIpan-4.0 (36), can provide reasonably accurate predictions for the binding core. Furthermore, the flexibility of PFRs poses additional hurdles. To address these challenges, the development of fast and pan-allelic pMHC-II modelling software is required to integrate prediction of the binding core and generation of plausible conformations for the entire peptide bound to MHC-II.

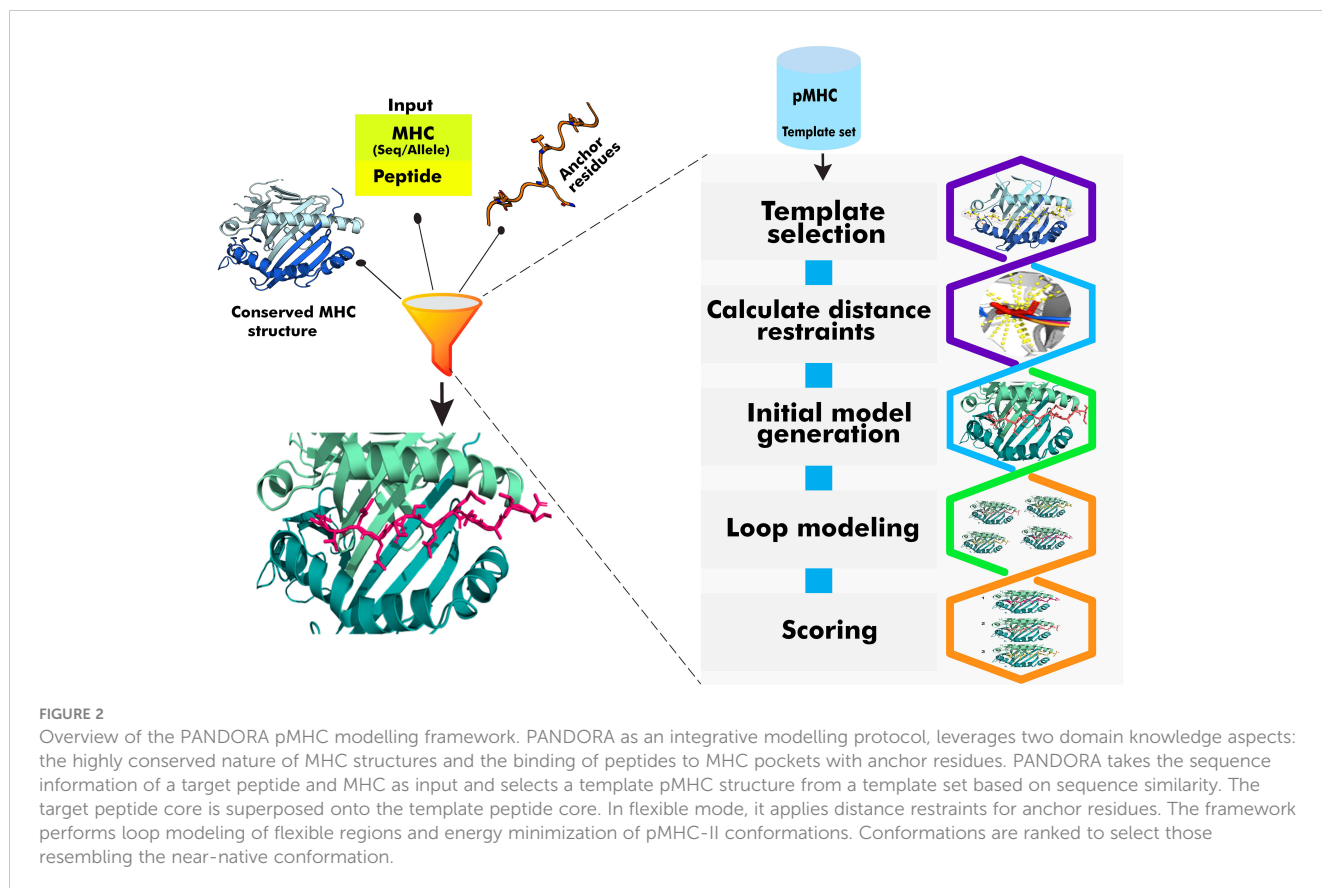
We present here the utility and performance of our pMHC modelling software, PANDORA v2.0, for pMHC-II modeling and its new version updates. We have earlier demonstrated PANDORA's reliable performance for modeling pMHC-I complexes (37, 38). PANDORA leverages two pieces of domain knowledge: 1) the high conservation of MHC structures and 2) the anchoring of peptides to the main pockets of MHC molecules (Figure 2). We benchmarked PANDORA on 136 experimentally resolved pMHC-II structures, including mouse alleles. When compared with an existing pMHC-II modelling technique, pDock (32), and also with AlphaFold (39), we show that PANDORA outperforms these methods in terms of generated model quality and computational efficiency. Additionally, we evaluate the effectiveness of the anchor prediction tool used in our approach (NetMHCIIpan-4.0). PANDORA's quality and speed show the potential for boosting structure-based Deep Learning (DL)

algorithms, making it a valuable tool in developing effective vaccine designs. We also discuss the existing limitations of anchor predictions and propose the integration of a structural and physics-based anchor predictor as a potential solution. Furthermore, we highlight the importance of further research in the modeling of post-translational modifications (PTMs) on peptide-MHC interactions.

2 Materials and methods

2.1 Structural template set building

Building the template dataset is similar to our previous work (37) and is expanded to make it suitable for pMHC-II. PANDORA retrieved structures of pMHC-II complexes from IMGT/3Dstructure-DB (40) and filters for those with peptides of lengths between 7 and 25 residues. Structures including the DM chaperone and the CLIP peptide, both known to affect the MHC-II conformation, are discarded (21, 41). The MHC-II alpha chain is renamed as chain "M" and the beta chain is renamed as chain "N" to make a distinction from MHC-I β 2-microglobulin which is renamed as chain "B". The peptide chain is renamed as chain "P". For the benchmark experiment presented in this work, the parsing resulted in a total of 136 pMHC-II templates, spanning over 32 α chain alleles, 81 β chain alleles, and a total of 44 unique MHC-II $\alpha\beta$ pairs (see details in Supplementary Table S1).



2.2 BLAST databases generation

BLAST (v2.10) is used to assign allele names (needed by NetMHCIIpan-4.0 for predicting binding cores) to the MHC sequences provided by the user and, independently, for the template selection step. The current version of PANDORA uses two BLAST databases. The first one (BLAST-DB1) is generated from the manually curated MHC sequences taken from <https://www.ebi.ac.uk/ipd/>, and it is used to assign the allele name to any MHC sequence provided by the user. This allele name will later be used as input for NetMHCIIpan4.0 to predict the binding core (see Template selection). The second one (BLAST-DB2) is generated from the template set sequences extracted by the PDB files retrieved as described above, and it is used for the template selection step.

2.3 Template selection

The template selection step has been updated from the first version of PANDORA (which used allele type names to identify templates) to a BLAST-based template selection. First, the target MHC sequences are queried against the BLAST-DB2 database with default parameters, and the results are ranked by percentage sequence identities. Templates sharing the highest sequence identity with the target sequences are selected and further ranked by peptide alignment score. Our peptide alignment method includes alignment of the binding core of the peptides followed by the addition of gaps at both their termini to account for different peptide lengths. The binding cores of the templates are derived from their corresponding structures. The binding core for the query peptide is predicted by NetMHCIIpan4.0. The peptides' alignments are then scored using a PAM30 substitution matrix. The highest-ranking template is then selected for modeling.

2.4 Modeling

We perform 3D modeling as described previously. For MHC-II, we restrain four anchor positions (P1, 4, 6, and 9) while keeping the peptide flanking regions flexible during the modeling step. In the default mode for pMHC-II cases, the whole peptide core is kept fixed as the template conformation. PANDORA v2.0 also supports restraints-flexible modelling mode for the peptide core, where users can provide anchors' restraints standard deviation, thereby specifying the extent of deviation of restraints from those in the templates in Angstroms. By default, 20 (adjustable) 3D models are produced, which are ranked by MODELLER's (42) internal molpdf score.

2.5 L-RMSD calculation

The L-RMSD is calculated as described in (43) as the backbone L-RMSD (including only the backbone atoms N, C α , C, and O). We calculate "Core L-RMSD" for the binding core residues of the

peptide, "Flanking L-RMSD" for the flanking regions of the peptide (i.e., the residues at the N-terminal of the first anchor and at the C-terminal of the fourth anchor), and "Whole L-RMSD" for all the residues of the peptide. The lower the L-RMSD, the better a model is.

3 Results

3.1 Modeling performance on the benchmark set

We benchmarked PANDORA's performance in reproducing X-ray crystal structures of pMHC-II complexes from the template set ($n = 136$). We carried out a leave-one-out validation approach where we iteratively removed a structure from the template database and allowed PANDORA to predict the pMHC-II complex using sequence and anchor information. To rule out the impact of anchor predictions, the anchor positions provided to PANDORA in this experiment were obtained from the target experimental structure to assess the modelling quality (see discussion on anchor prediction effects in the "NetMHCIIpan's anchor prediction" section).

We analyzed the distribution of the best model (i.e., the model with the lowest L-RMSD) conformations obtained for the whole and core peptide regions (Figures 3A, C, E; detailed information on different RMSD values is reported in Supplementary Table S2). The results demonstrate that for 91.1% (125 out of 136) cases, PANDORA was able to sample at least one high-quality model (whole peptide L-RMSD $< 2 \text{ \AA}$) with an overall mean L-RMSD of $1.11 \pm 0.86 \text{ \AA}$ (i.e., Figure 3B). A small number of cases (11 out of 136) showed a relatively higher whole peptide L-RMSD of $> 2 \text{ \AA}$ (see Figures 3E, F, and "The PFR Conformation Evaluation" section). We investigated the distribution of whole and core L-RMSDs over various peptide lengths, as illustrated in Figure 3A. Our analysis reveals a correlation between peptide lengths and the L-RMSD values, with longer peptides exhibiting higher L-RMSD values (Supplementary Figures S1A, B).

Furthermore, in terms of model ranking, we examined the performance of PANDORA by reporting L-RMSD for the top-ranked model (i.e., the conformation ranked as the top model using molpdf scoring function) (Figures 3D, F; for details, see Supplementary Figure S2). Our results show that PANDORA achieved an 85% success rate (L-RMSD $< 2 \text{ \AA}$) for the top 5 ranked models in the entire template set.

3.2 PANDORA generates low-energy conformations for the binding core

With four anchor positions in the binding groove, the structure of pMHC-II is well-suited for a restraint-based modelling approach. With the default mode (see Modelings in Methods), PANDORA demonstrates high accuracy in reproducing high-quality core conformations, with an average core L-RMSD of $0.49 \pm 0.27 \text{ \AA}$

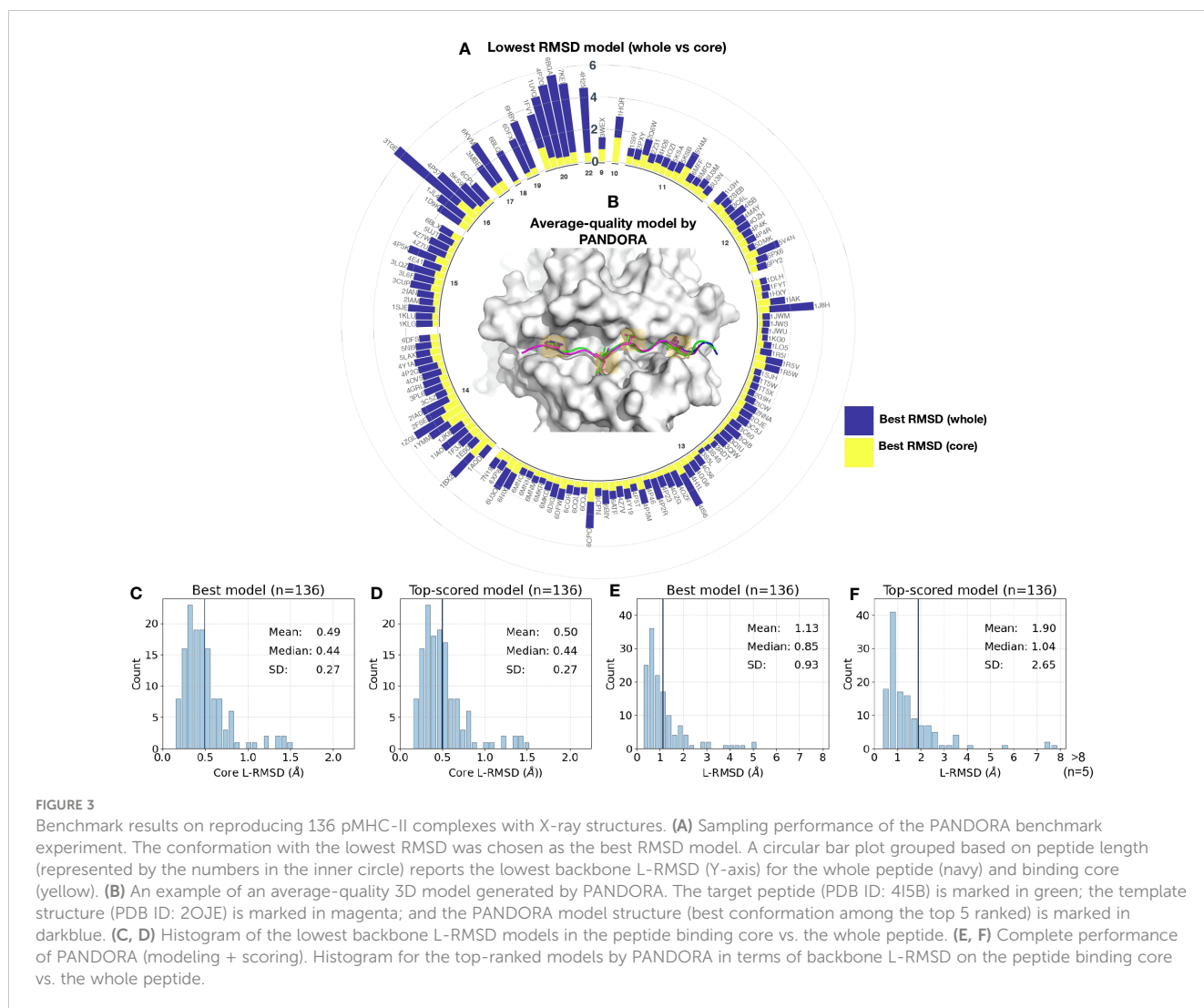


FIGURE 3

Benchmark results on reproducing 136 pMHC-II complexes with X-ray structures. **(A)** Sampling performance of the PANDORA benchmark experiment. The conformation with the lowest RMSD was chosen as the best RMSD model. A circular bar plot grouped based on peptide length (represented by the numbers in the inner circle) reports the lowest backbone L-RMSD (Y-axis) for the whole peptide (navy) and binding core (yellow). **(B)** An example of an average-quality 3D model generated by PANDORA. The target peptide (PDB ID: 4I5B) is marked in green; the template structure (PDB ID: 2OJE) is marked in magenta; and the PANDORA model structure (best conformation among the top 5 ranked) is marked in darkblue. **(C, D)** Histogram of the lowest backbone L-RMSD models in the peptide binding core vs. the whole peptide. **(E, F)** Complete performance of PANDORA (modeling + scoring). Histogram for the top-ranked models by PANDORA in terms of backbone L-RMSD on the peptide binding core vs. the whole peptide.

(93.38% of the cases having an L-RMSD < 1 Å) (Figures 3C, E). The fully-flexible mode, which allows for flexibility in the peptides' binding core, yielded an average core L-RMSD of 0.47 ± 0.2 Å (Supplementary Figure S3). However, the restraints-flexible mode increases the computational time by 90%, while marginally enhancing the overall quality (~ 6.46 min/case in the fully flexible mode vs. 3.75 min/case in the default mode).

3.3 Comparisons with AlphaFold and pDock

We compared PANDORA's performance against existing approaches, such as pDOCK (32) and AlphaFold (39). To assess the general performance of the pipeline, we used NetMHCIIpan's predicted anchor positions for this comparison.

pDock uses the ICM (Internal Coordinate Mechanics) algorithm to perform a flexible peptide docking into the MHC binding groove. During docking, the position of the peptide is only loosely constrained so that it retains a conformation close to its initial structure. For comparisons against pDock, we modeled pMHC-II complexes using PANDORA for the cases reported by

Khan and Ranganathan (32). We obtained a mean L-RMSD of 0.27 ± 0.07 Å for C α core while pDock achieved 0.59 ± 0.24 Å (Table 1). pDock retained RMSD estimates by redocking experimental pMHC X-ray structures; thus, the core residues are referred to a priori. PANDORA automatically predicts anchor residues (using NetMHCIIpan-4.0 (36)) and a suitable template, generating higher-quality peptide core conformations. We did not use pDock to perform cross-docking on our template set since pDock is not publicly available for download and usage.

We also compared PANDORA with one of the best AI methods available for protein structure predictions, i.e., AlphaFold. AlphaFold is an advanced deep neural network approach that achieves unprecedented accuracy in protein folding predictions (44). However, since AlphaFold relies on sequence conservation information, it performs poorly on proteins where such information is absent, such as antibody-antigens and peptides (e.g., synthetic peptides or frame-shift mutated peptides) (45). For an objective comparison, we chose to use a version of AlphaFold that also uses templates to predict MHC structures (colabfold (44)). Our comparison shows that not all AlphaFold-generated pMHC-II conformations have the correct anchor positions. Out of four randomly selected cases (Figures 4A–D), in two cases (PDB ID:

TABLE 1 Comparison of PANDORA and pDock in pMHC-II modelling.

PDB	PANDORA's best core C α L-RMSD (Å)	pDOCK C α core L-RMSD(Å)*
1FYT	0.38	0.35
1KLU	0.30	0.59
1T5W	0.24	0.65
1PYW	NA	0.32
1SJE	0.21	0.37
1AQD	0.24	1.01

NA, not available. At this stage, PANDORA could not model one case as the peptide sequence includes two non-canonical residues not handled by MODELLER.

The calculated core C α L-RMSD (Å) on modeling 5 pMHC-II complexes using integrative homology modeling and a grid-based docking method. PANDORA's best model quality is compared to pDock as no pDock scoring function was disclosed in pDock so it seems that pDock reported the best RMSDs in their paper. *Data extracted from Khan & Ranganathan (32).

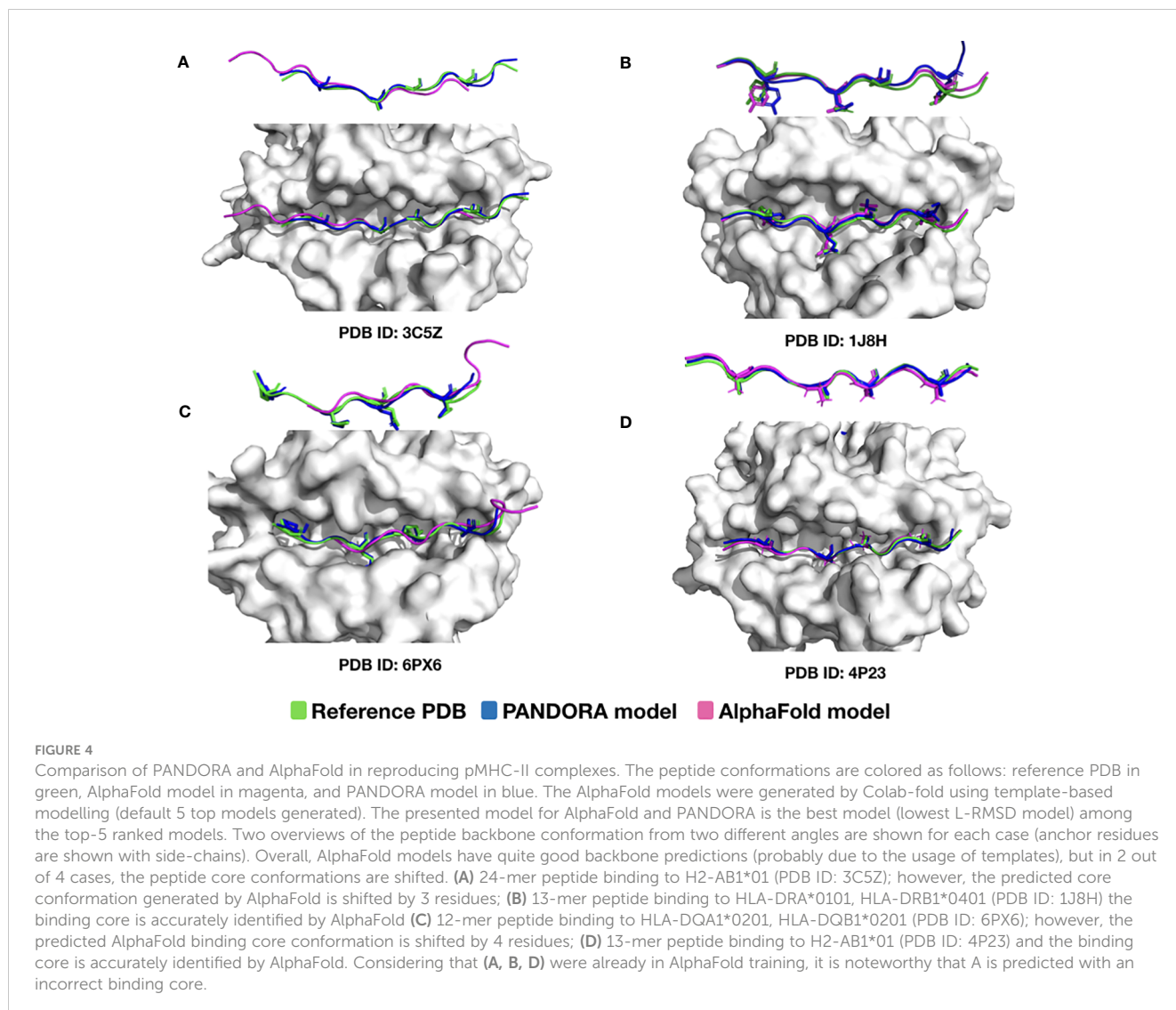
3C5Z and 6PX6), AlphaFold was unsuccessful in predicting the peptide's conformation with the correct anchor residues (Figures 4A, C). Notably, they were both part of AlphaFold's training set. This is mainly because PANDORA correctly identified

the binding core for the four cases using NetMHCIIpan binding core predictions (see "NetMHCIIpan4.0 performance" in the next section).

Additionally, considering computational cost, PANDORA outperforms AlphaFold (regarding resources) and pDock. PANDORA is much more efficient considering template selection, anchor prediction, and modeling require ~3-4 minutes (from 3.75 to 6.46 minutes per case, depending on the mode, with shorter times for pMHC-I) on one core from an Intel(R) Xeon(R) Gold 6142 CPU @ 2.60 GHz. While pDock reported requiring 10 minutes for modeling on 2 CPUs 3.20 GHz (without homology modelling). Also, AlphaFold requires a significant amount of computation power-up to 18 GB of GPU power and 20 minutes to model a single pMHC case.

3.4 The impacts of binding core prediction on PANDORA model quality

The interaction between the peptide binding core and the MHC binding groove directly impacts the quality of a model; therefore, choosing the correct binding core is critical. In the absence of user-



defined anchor residues, PANDORA uses NetMHCIIpan to predict the binding core. Hence, we evaluate NetMHCIIpan's binding core prediction accuracy by comparing its predictions to known cores from experimental PDB structures. Our results show that the anchors were incorrectly predicted in 33 of the 136 cases in the benchmark dataset. In most cases, the observed shifts were by one or two residues (26 of 33), but misalignments of up to 8 residues were also observed (Supplementary Figure S4).

3.5 PANDORA as a pan-allele modelling method

Owing to the high structural similarity across MHC-II alleles, it is possible to model pMHC-II complexes using different MHC-II alleles as templates. Our results show that even when a template with the same MHC allele type for either of the chains was not available in the template set (25% of cases), PANDORA was still able to provide models with a mean L-RMSD of 0.86 Å for the best-RMSD models and 1.05 Å for the top 10 ranked models (Figure 5).

3.6 Software improvements

PANDORA v2.0 includes major improvements from the first release:

Frontend (User side):

- Capability to use MHC-sequence as input instead of only allele name, leading to much broader allele coverage than version 1.0.
- Addition of command-line interface for easier accessibility and bash integration.
- Addition of restraints-flexible modelling mode to avoid small clashes caused by rigid restraints (see Materials and Methods).
- Improvements in the python user interface.
- Easier software and database installation.
- Addition of an option to remove or keep beta2-microglobulin in the generated models, as Beta2-microglobulin can be crucial or not, depending on what the models will be used for (MD, AI, manual exploration, etc.).

Backend (internal software side):

- BLAST-based template selection instead of allele-name based template selection.
- Addition of a reference sequence database for allele names and MHC sequence automatic retrieval.
- The allele name is now automatically retrieved with BLAST when only the sequence is provided.
- Improvement in the MHC-II template parsing to prevent multiple structures from being discarded or from missing the allele name.
- Addition of parallelization and minor optimization improvements for the template set generation, drastically increasing its speed.

4 Discussion

PANDORA is a 3D modelling software for both pMHC-I and -II. Here we evaluated the performance of PANDORA on reliably generating peptide conformations binding to MHC class II complexes alongside the software improvements. We applied homology-driven restraint-based modelling to reduce the computational time during sampling (3.75 min/case on one CPU core). The proposed method was tested on 136 complexes, making it the largest modeling effort of pMHC-II complexes to date. Our results show that PANDORA was able to effectively model these complexes, achieving an 85% success rate (L-RMSD < 2 Å) for the top 5 ranked models in the entire template set and generating particularly high-quality peptide core conformations.

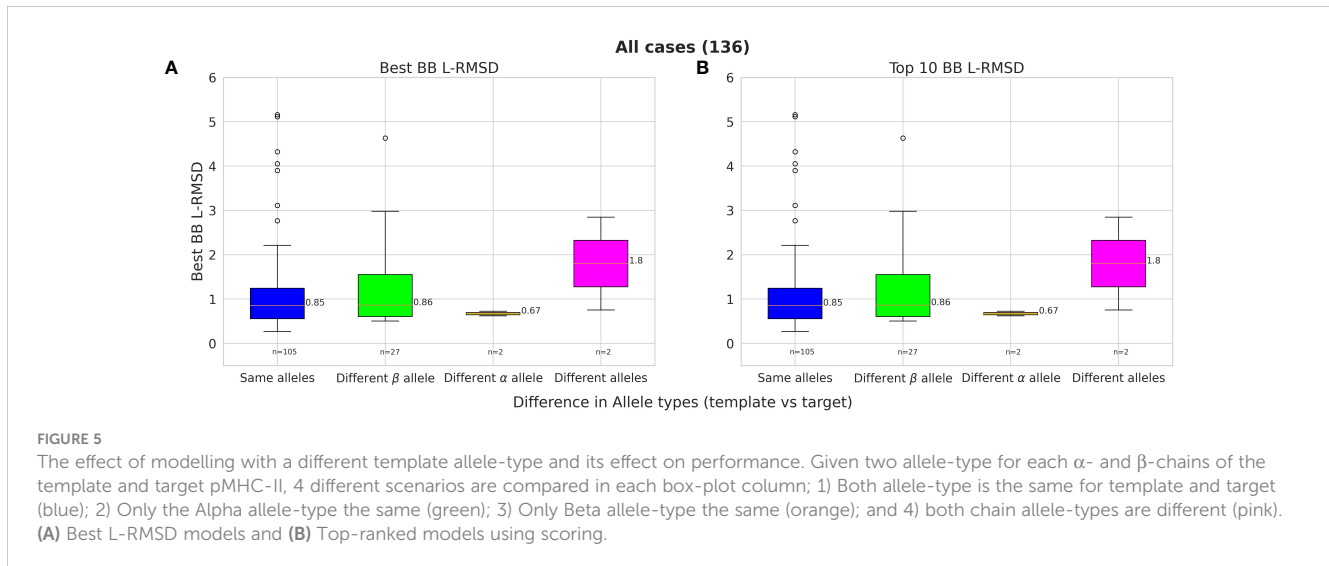
PANDORA outperforms pDock (32) and AlphaFold (39) regarding computational time and core L-RMSD values. PANDORA incorporates domain knowledge into the modeling. In contrast, AlphaFold is a general protein structure prediction method that relies on sequence conservation information, and conservation on the peptide side has little or no bearing on this binding. Our comparison shows that not all AlphaFold-generated pMHC-II conformations have the correct anchor positions. AlphaFold's higher computational cost is a major impediment to model millions of pMHCs, whereas PANDORA is a more practical choice.

PANDORA has the following unique features: 1) Fast: enabling high-throughput modeling of 3D pMHC-II complexes; 2) Reliable: generating low-energy models; 3) Efficient: With the use of anchor distance restraints, it to work on both MHC-I and MHC-II; 4) Template availability: providing an extensively cleaned template database of pMHC complexes, valuable for reliable homology modeling; 5) Highly Modular: It is easy to customize or extend; 6) Pan-allele: User may include MHCs from different species.

PANDORA has a user-friendly interface allowing users to incorporate new configurations such as 1) more extensive sampling (especially with longer peptides); 2) specification of secondary structure restraints (23% of benchmark cases formed beta-strand PFR, Supplementary Figure S5D); 3) fully fixed mode vs. flexible mode for the core conformation; 4) Manually defining the anchor residues; 5) Possibility of changing to other anchor predictor software. Its highly modular framework (Supplementary Figure S7) facilitates future community-wide development.

Knowledge of the peptide binding core is required to generate the pMHC-II complex structure. When the user doesn't input the anchor residues' position, PANDORA currently relies on NetMHCIIpan-4.0 as an anchor predictor (36). This software has a limited, yet large, set of available MHC alleles to utilize, and it can sometimes fail to predict the correct binding core (Supplementary Figure S4). Using an anchor predictor relying on structural and physics-based data could overcome these limitations for the pMHC anchor prediction, allowing for more accurate, pan-allelic anchor predictions.

PFR can influence TCR interactions (46–49); introducing a modeling program to generate credible PFR conformations is an important step forward. It is important to note that a singular X-ray structure exclusively depicts only one snapshot of the complex conformation. This implies that a method could generate a possible



PFR conformation that is not currently cataloged in the PDB but holds biological significance. To address this issue, PANDORA generates an ensemble of near-native conformations (top N-ranked conformations).

Further work is needed to model the post-translational modifications (PTM) in peptides binding to MHC, which have been shown to modulate antigen presentation and recognition (50, 51) and, moreover, PTMs on peptides increase the vast number of possible pMHC combinations. PTMs have a structural impact on the stability of pMHC complexes and the consequent modulations of immune responses (52). Although it has not yet been extensively evaluated within our framework, we recognize its potential benefits for the field and remain committed to conducting additional research and possibly incorporating this method into our future research.

While PANDORA provides energy scores, its primary focus is on 3D modeling rather than predicting binding affinity, it might be possible to utilize the energy scores from PANDORA models or running molecular dynamics on PANDORA models to gain insights into MHC binding specificities. In addition, PANDORA can potentially contribute to advancing our understanding of cancer biology, particularly in unraveling the impact of peptide mutations on MHC binding and the exposure of peptide side chains to T-cells or (see * marked cases in [Supplementary Tables S1, S2](#)). Although not intended for neoantigen identification, PANDORA was used to evaluate the effects of point mutations on a melanoma tumor antigen. PANDORA accurately modeled both peptides' side chains (see [Supplementary Figure S6](#)), resulting in high-affinity energy scores and a slight improvement in mutant binding.

In conclusion, the ability of PANDORA to generate high-quality peptide conformations within the MHC-II binding groove lends great reliability to the models employed for analyzing molecular interactions at the atomic level. Due to PANDORA's computational efficiency, initial conformations for molecular dynamics simulations can be quickly built.

It is now feasible to enrich the actively accumulating wealth of pMHC binding affinity and mass spectrometry data with physics-based PANDORA models and aid structure-boosted artificial intelligence algorithms in identifying antigenic peptides (for example,

by training the deep learning framework DeepRank on these 3D models). As such, it can be leveraged to identify cancer neoantigens or viral antigenic peptides that hold promise as vaccine candidates. It will therefore pave the way for developing novel cancer immunotherapies.

Data availability statement

The template database and package are available at DOI: 10.5281/zenodo.6373630 PANDORA is available as a Python package on conda at <https://anaconda.org/csb-nijmegen/csbpandora>. The package source code and the documentation are available on GitHub at <https://github.com/X-lab3D/PANDORA>. Further inquiries can be directed to the corresponding authors.

Author contributions

FP: Conceptualization, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. DM: Conceptualization, Data curation, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. GR: Formal Analysis, Validation, Visualization, Writing – review & editing. PH: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. MK-J: Formal Analysis, Supervision, Visualization, Writing – review & editing. LX: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. LX, FP, and DM received support from the Hypatia Fellowship from Radboudumc (Rv819.52706). GR is supported by Europees Fonds voor Regionale Ontwikkeling (EFRO) (R0005582).

Acknowledgments

The authors would like to thank Giulia Crocioni, Derek van Tilborg, and Shahiel Maassen for their help with the software development. We also thank Hassan Rasouli and Yannick J.M. Aarts for their assistance in plot visualization.

Conflict of interest

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

References

- Janeway C. *Immunobiology: the immune system in health and disease*. 5. ed. New York, NY: Garland Publ (2001). p. 732.
- Blum JS, Wearsch PA, Cresswell P. Pathways of antigen processing. *Annu Rev Immunol* (2013) 31(1):443–73. doi: 10.1146/annurev-immunol-032712-095910
- Madden DR. The three-dimensional structure of peptide-MHC complexes. *Annu Rev Immunol* (1995) 13(1):587–622. doi: 10.1146/annurev-iy.13.040195.003103
- Abualrous ET, Stolzenberg S, Sticht J, Wiczorek M, Roske Y, Günther M, et al. MHC-II dynamics are maintained in HLA-DR allotypes to ensure catalyzed peptide exchange. *Nat Chem Biol* (2023) 19:1–9. doi: 10.1038/s41589-023-01316-3
- Thomas C, Tampé R. Structure of the TAPBPR-MHC I complex defines the mechanism of peptide loading and editing. *Science* (2017) 358(6366):1060–4. doi: 10.1126/science.aao6001
- Zhang C, Anderson A, DeLisi C. Structural principles that govern the peptide-binding motifs of class I MHC molecules. *JMB* (1998) 281(5):929–47. doi: 10.1006/jmbi.1998.1982
- Sinigaglia F, Hammer J. Defining rules for the peptide-MHC class II interaction. *Curr Opin Immunol* (1994) 6(1):52–6. doi: 10.1016/0952-7915(94)90033-7
- Ferrante A, Gorski J. Cooperativity of hydrophobic anchor interactions: evidence for epitope selection by MHC class II as a folding process. *J Immunol* (2007) 178(11):7181–9. doi: 10.4049/jimmunol.178.11.7181
- Murthy VL, Stern LJ. The class II MHC protein HLA-DR1 in complex with an endogenous peptide: implications for the structural basis of the specificity of peptide binding. *Structure* (1997) 5(10):1385–96. doi: 10.1016/S0969-2126(97)00288-8
- Wucherpfennig KW. Insights into autoimmunity gained from structural analysis of MHC-peptide complexes. *Curr Opin Immunol* (2001) 13(6):650–6. doi: 10.1016/S0952-7915(01)00274-6
- Madura F, Rizkallah PJ, Holland CJ, Fuller A, Bulek A, Godkin AJ, et al. Structural basis for ineffective T-cell responses to MHC anchor residue-improved “heteroclitic” peptides: Molecular immunology. *Eur J Immunol* (2015) 45(2):584–91. doi: 10.1002/eji.201445114
- Smith AR, Alonso JA, Ayres CM, Singh NK, Hellman LM, Baker BM. Structurally silent peptide anchor modifications allosterically modulate T cell recognition in a receptor-dependent manner. *Proc Natl Acad Sci USA* (2021) 118(4):e2018125118. doi: 10.1073/pnas.2018125118
- Saotome K, Dudgeon D, Colotti K, Moore MJ, Jones J, Zhou Y, et al. Structural analysis of cancer-relevant TCR-CD3 and peptide-MHC complexes by cryoEM. *Nat Commun* (2023) 14(1):2401. doi: 10.1038/s41467-023-37532-7
- Sušac L, Vuong MT, Thomas C, Von Bülow S, O'Brien-Ball C, Santos AM, et al. Structure of a fully assembled tumor-specific T cell receptor ligated by pMHC. *Cell* (2022) 185(17):3201–3213.e19. doi: 10.1016/j.cell.2022.07.010
- Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* (2017) 547(7662):222–6. doi: 10.1038/nature23003
- Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* (2017) 547(7662):217–21. doi: 10.1038/nature22991
- Kreiter S, Vormehr M, van de Roemer N, Diken M, Löwer M, Diekmann J, et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature* (2015) 520(7549):692–6. doi: 10.1038/nature14426

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/fimmu.2023.1285899/full#supplementary-material>

- Alspach E, Lussier DM, Miceli AP, Kizhvatov I, DuPage M, Luoma AM, et al. MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* (2019) 574(7780):696–701. doi: 10.1038/s41586-019-1671-8
- Tomasello G, Armenia I, Molla G. The Protein Imager: a full-featured online molecular viewer interface with server-side HQ-rendering capabilities. *Bioinformatics* (2020) 36(9):2909–11. doi: 10.1093/bioinformatics/btaa009
- Jardetzky TS, Brown JH, Gorga JC, Stern LJ, Urban RG, Strominger JL, et al. Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline II-like conformation for bound peptides. *Proc Natl Acad Sci USA* (1996) 93(2):734–8. doi: 10.1073/pnas.93.2.734
- Wiczorek M, Abualrous ET, Sticht J, Álvaro-Benito M, Stolzenberg S, Noé F, et al. Major histocompatibility complex (MHC) class I and MHC class II proteins: conformational plasticity in antigen presentation. *Front Immunol* (2017) 8:292. doi: 10.3389/fimmu.2017.00292
- Hammer J, Bono E, Gallazzi F, Belunis C, Nagy Z, Sinigaglia F. Precise prediction of major histocompatibility complex class II-peptide interaction based on peptide side chain scanning. *J Exp Med* (1994) 180(6):2353–8. doi: 10.1084/jem.180.6.2353
- Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, et al. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* (1994) 368(6468):215–21. doi: 10.1038/368215a0
- Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, et al. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* (1993) 364(6432):33–9. doi: 10.1038/364033a0
- Kulski JK, Suzuki S, Shiina T. Human leukocyte antigen super-locus: nexus of genomic supergenes, SNPs, indels, transcripts, and haplotypes. *Hum Genome Var* (2022) 9(1):49. doi: 10.1038/s41439-022-00226-5
- Barker DJ, Maccari G, Georgiou X, Cooper MA, Flicek P, Robinson J, et al. The IPD-IMGT/HLA database. *Nucleic Acids Res* (2023) 51(D1):D1053–60. doi: 10.1093/nar/gkac1011
- Berman HM. The protein data bank. *Nucleic Acids Res* (2000) 28(1):235–42. doi: 10.1093/nar/28.1.235
- Patronov A, Dimitrov I, Flower DR, Doytchinova I. Peptide binding prediction for the human class II MHC allele HLA-DP2: a molecular docking approach. *BMC Struct Biol* (2011) 11(1):32. doi: 10.1186/1472-6807-11-32
- Patronov A, Salamanova E, Dimitrov I, Flower D, Doytchinova I. Histidine Hydrogen Bonding in MHC at pH 5 and pH 7 Modeled by Molecular Docking and Molecular Dynamics Simulations. *CAD* (2014) 10(1):41–9. doi: 10.2174/15734099113096660050
- Bordner AJ. Towards universal structure-based prediction of class II MHC epitopes for diverse allotypes. *PLoS One* (2010) 5(12):e14383. doi: 10.1371/journal.pone.0014383
- Tong JC, Tan TW, Ranganathan S. Modeling the structure of bound peptide ligands to major histocompatibility complex. *Protein Sci* (2004) 13(9):2523–32. doi: 10.1110/ps.04631204
- Khan J, Ranganathan S. pDOCK: a new technique for rapid and accurate docking of peptide ligands to Major Histocompatibility Complexes. *Immunome Res* (2010) 6. doi: 10.1186/1745-7580-6-S1-S2
- Atanasova M, Patronov A, Dimitrov I, Flower DR, Doytchinova I. EpiDOCK: a molecular docking-based tool for MHC class II binding prediction. *PEDS* (2013) 26(10):631–4. doi: 10.1093/protein/gz018
- Racle J, Michaux J, Rockinger GA, Arnaud M, Bobisse S, Chong C, et al. Robust

prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. *Nat Biotechnol* (2019) 37(11):1283–6. doi: 10.1038/s41587-019-0289-6

35. Liu Z, Jin J, Cui Y, Xiong Z, Nasiri A, Zhao Y, et al. DeepSeqPanII: an interpretable recurrent neural network model with attention mechanism for peptide-HLA class II binding prediction. *IEEE/ACM Trans Comput Biol Bioinf.* (2022) 19(4):2188–96. doi: 10.1109/TCBB.2021.3074927
36. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. 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. *NAR* (2020) 48(W1):W449–54. doi: 10.1093/nar/gkaa379
37. Marzella DF, Parizi FM, Tilborg Dv, Renaud N, Sybrandi D, Buzatu R, et al. PANDORA: A fast, anchor-restrained modelling protocol for peptide: MHC complexes. *Front Immunol* (2022) 13:878762. doi: 10.3389/fimmu.2022.878762
38. Marzella DF, Crocioni G, Parizi FM, Xue LC. The PANDORA software for anchor-restrained peptide:MHC modeling. *Methods Mol Biol* (2023) 2673:251–71. doi: 10.1007/978-1-0716-3239-0_18
39. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* (2021) 596:1–11. doi: 10.1038/s41586-021-03819-2
40. Ehrenmann F, Kaas Q, Lefranc MP. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. *Nucleic Acids Res* (2010) 38:D301–7. doi: 10.1093/nar/gkp946
41. Neeffes J, Jongsma MLM, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* (2011) 11(12):823–36. doi: 10.1038/nri3084
42. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinf* (2016) 54(5):1–5. doi: 10.1002/prot.21804
43. Lensink MF, Méndez R, Wodak SJ. Docking and scoring protein complexes: CAPRI 3rd Edition. *Proteins* (2007) 69(4):704–18. doi: 10.1002/prot.21804
44. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat Methods* (2022) 19(6):679–82. doi: 10.1038/s41592-022-01488-1
45. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. *Bioinformatics*; (2021) 4:2021–10. doi: 10.1101/2021.10.04.463034
46. Carson RT, Vignali KM, Woodland DL, Vignali DAA. T cell receptor recognition of MHC class II-bound peptide flanking residues enhances immunogenicity and results in altered TCR V region usage. *Immunity* (1997) 7(3):387–99. doi: 10.1016/S1074-7613(00)80360-X
47. Vignali DA, Strominger JL. Amino acid residues that flank core peptide epitopes and the extracellular domains of CD4 modulate differential signaling through the T cell receptor. *JEM* (1994) 179(6):1945–56. doi: 10.1084/jem.179.6.1945
48. Muller CP, Ammerlaan W, Fleckenstein B, Krauss S, Kalbacher H, Schneider F, et al. Activation of T cells by the ragged tail of MHC class II-presented peptides of the measles virus fusion protein. *Int Immunol* (1996) 8(4):445–56. doi: 10.1093/intimm/8.4.445
49. Zavala-Ruiz Z, Strug I, Walker BD, Norris PJ, Stern LJ. A hairpin turn in a class II MHC-bound peptide orients residues outside the binding groove for T cell recognition. *Proc Natl Acad Sci USA* (2004) 101(36):13279–84. doi: 10.1073/pnas.0403371101
50. Levy R, Alter Regev T, Paes W, Gumpert N, Cohen Shvefel S, Bartok O, et al. Large-scale immunopeptidome analysis reveals recurrent posttranslational splicing of cancer- and immune-associated genes. *MCP* (2023) 22(4):100519. doi: 10.1016/j.mcpro.2023.100519
51. Bloodworth N, Barbaro NR, Moretti R, Harrison DG, Meiler J. Rosetta FlexPepDock to predict peptide-MHC binding: An approach for non-canonical amino acids. *PLoS One* (2022) 17(12):e0275759. doi: 10.1371/journal.pone.0275759
52. Sandalova T, Sala BM, Achour A. Structural aspects of chemical modifications in the MHC-restricted immunopeptidome; Implications for immune recognition. *Front Chem* (2022) 10:861609. doi: 10.3389/fchem.2022.861609