HLA-Arena: A customizable environment for the structural modeling and analysis of peptide-HLA complexes for cancer immunotherapy

Dinler A. Antunes, PhD[*1], Jayvee R. Abella[*1], Sarah Hall-Swan[1], Didier Devaurs, PhD[2], Anja Conev[1], Mark Moll, PhD[1], Gregory Lizée, PhD[3], and Lydia E. Kavraki, PhD[†,1]

[1]Department of Computer Science, Rice University, Houston, TX 77005, USA
[2]Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP[‡], LJK, 38000 Grenoble, France
[3]Department of Melanoma Medical Oncology - Research, The University of Texas MD Anderson Cancer Center, Houston, TX 77054, USA

KEY POINT SUMMARY

Key objective: Enabling large-scale structural modeling and analysis of peptide-HLA complexes for cancer immunotherapy applications.
Knowledge generated: We created a customizable environment, called HLA-Arena, with user-friendly computational workflows that allow for varied structure-based analyses of peptide-HLA complexes. To illustrate this, we show how researchers can use HLA-Arena to perform geometry prediction of peptide binding modes, peptide binding energy prediction, and structure-based virtual screening of tumor-derived peptides, for any classical class I HLA of interest.
Relevance: HLA-Arena can be integrated in computational pipelines to support basic cancer research or to help inform physicians in pre-clinical settings. It can be used to perform structure-based selection of peptides for T-cell-based immunotherapy, neoantigen discovery, and vaccine development.

Keywords: Human leukocyte antigen (HLA); Peptide-HLA complex; Homology modeling; Molecular docking; Structural modeling; Structural analysis; Cancer immunotherapy

[*]These authors contributed equally to this work
[†]Corresponding author: kavraki@rice.edu, Rice University, MS132, P.O. Box 1892, Houston, TX 77251-1892, USA
[‡]Institute of Engineering Univ. Grenoble Alpes

ABSTRACT

Purpose Human Leukocyte Antigen (HLA) protein receptors play a key role in cellular immunity. They bind intracellular peptides and display them for recognition by T-cell lymphocytes. As T-cell activation is partially driven by structural features of these peptide-HLA complexes, their structural modeling and analysis is becoming a central component of cancer immunotherapy projects. Unfortunately, this kind of analysis is limited by the small number of experimentally-determined structures of peptide-HLA complexes. Overcoming this limitation requires developing novel computational methods to model and analyze peptide-HLA structures.

Design Here we describe a new platform for the structural modeling and analysis of peptide-HLA complexes, called HLA-Arena, that we have implemented using Jupyter Notebook and Docker. It is a customizable environment that facilitates the usage of computational tools, such as APE-Gen and DINC, that we previously applied to peptide-HLA complexes. By integrating other commonly-used tools, such as Modeller and MHCflurry, this environment includes support for diverse tasks in structural modeling, analysis, and visualization.

Results To illustrate the capabilities of HLA-Arena, we describe three example workflows applied to peptide-HLA complexes. Leveraging the strengths of our tools, DINC and APE-Gen, the first two workflows show how to perform geometry prediction for peptide-HLA complexes and structure-based binding prediction, respectively. The third workflow presents an example of large-scale virtual screening of peptides for multiple HLA alleles.

Conclusion These workflows illustrate the potential benefits of HLA-Arena for the structural modeling and analysis of peptide-HLA complexes. As HLA-Arena can easily be integrated within larger computational pipelines, we expect its potential impact to vastly increase. For instance, it could be used to conduct structural analyses for personalized cancer immunotherapy, neoantigen discovery or vaccine development.

Introduction

Immunotherapy treatments are now at the forefront of methods used for cancer therapy. These treatments aim at harvesting a patient's own immunological defenses to identify and eliminate cancer cells.[1] Many of these immunotherapy treatments involve class I human leukocyte antigen (HLA) protein receptors. HLA receptors bind peptides produced by the cleavage of intracellular proteins, which is a continuous process present in almost every cell. The resulting peptide-HLA (pHLA) complexes are then exposed at the surface of cells. Being also present in cancer cells, this mechanism allows circulating T-cell lymphocytes to recognize tumor-associated peptides, thus triggering T-cell activation, tumor elimination, and immunological memory against the tumor.[1,2]

It has been shown that immunological outcomes are partially driven by structural features of pHLA complexes.[2–4] Therefore, the structural modeling and analysis of these complexes is becoming essential to ensure the efficacy and safety of immunotherapy treatments.[2] However, pHLA structural features are impacted by the genetic variability of both patients and tumors.[2,5] First, the set of peptides available for presentation reflects the patient's genetic background and cancer-specific alterations.[2,5] Second, each individual has up to six class I HLA alleles,[6] among the nearly 19,000 alleles in the human population.[7] Each allele encodes for a receptor with specific characteristics, that will display a different pool of peptides. Therefore, the structural modeling and analysis of pHLA complexes for cancer immunotherapy requires fast and customizable methods that can handle patient-specific data.

Unfortunately, the cost and time requirements of gold-standard experimental techniques in structural biology prevent their use in personalized medicine. In addition, very few structures of pHLA complexes have been determined experimentally. Therefore, researchers have turned toward computational methods for the structural modeling of pHLA complexes. However, the length and flexibility of displayed peptides represent a major challenge for traditional methods.[5] As an alternative, in previous work we have developed several computational tools for the accurate and efficient modeling of pHLA complexes. For example, we have described a very fast method, called APE-Gen, to generate ensembles of peptide conformations bound to a given HLA receptor.[8] We have also developed a meta-docking approach, called DINC, which allows predicting binding modes of pHLA complexes.[9,10]

In this paper, we present a higher-level platform, called HLA-Arena, that allows carrying out a sophisticated structural modeling and analysis of pHLA complexes. Instead of having to deal with several computational tools, HLA-Arena provides researchers with a single customizable environment that fully integrates the tools we have developed, as well as other commonly-used software. HLA-Arena simplifies the interactions with these tools by leveraging the capabilities of Jupyter Notebook and Docker. It allows users to perform various workflows, each one involving a specific combination of tools and steps within a coherent scenario. Besides APE-Gen and DINC, HLA-Arena currently integrates Modeller,[11] for homology modeling, MHCflurry,[12] for binding affinity prediction, and NGL Viewer,[13] for structure visualization, among others.

Here, we present three example workflows illustrating the capabilities of HLA-Arena. The first one relies on DINC to predict the binding modes of two known peptides with their corresponding HLA receptors (i.e., *geometry prediction*). The second workflow relies on APE-Gen to assess differences in binding between peptides restricted to a given HLA receptor, based on generated binding mode ensembles (i.e., *binding prediction*). The third workflow aims at performing structure-based *virtual screening*, which requires speed and scalability. Using real immunopeptidomics data and a fictitious diplotype (i.e., six classical class I HLA alleles) we show how MHCflurry and APE-Gen can complement each other to select target-peptides for a hypothetical immunotherapy treatment.

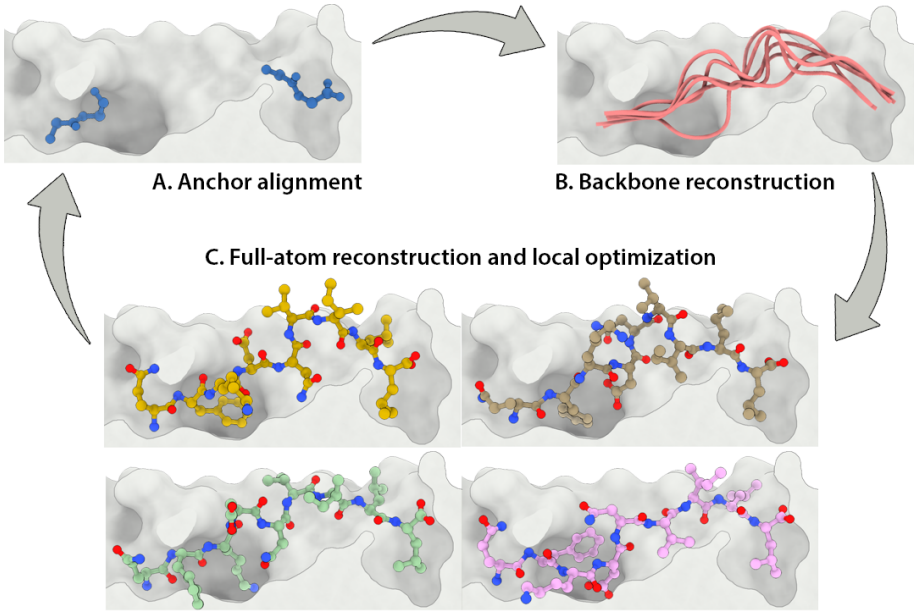


Figure 1. Generating binding mode ensembles with APE-Gen. A. Templates of backbone termini are used to position the anchor residues of a peptide in the binding site. **B.** The random coordinate descent (RCD) loop-closure tool[19] is utilized to generate an ensemble of backbone conformations for this peptide. **C.** Full-atom reconstruction of peptide side-chains and local optimization of the resulting complex are performed for each sampled backbone. The highest-quality binding mode can be selected to be used as template for the next round of the iterative process.

Design

Computational approaches for pHLA binding mode prediction

Despite their huge sequence diversity, HLA receptors feature very conserved secondary and tertiary structures, as illustrated by available data.[14–16] Such conserved folding makes HLA modeling an easy task with tools leveraging homology modeling.[8,17,18] On the other hand, predicting the binding modes of peptides to HLA receptors is much harder because of the size and flexibility of these peptides. As recently reviewed, strategies used to overcome this challenge include (i) constrained backbone prediction, (ii) constrained termini prediction, and (iii) incremental prediction.[5]

In recent years, we have implemented two computational approaches for pHLA binding mode prediction using these strategies. The first one, called APE-Gen (anchored peptide-MHC ensemble generator), can quickly produce an ensemble of binding modes for a pHLA complex, using termini templates to position the peptide in the HLA binding cleft (*see* Fig. 1 and Supplemental Material).[8] The second one, called DINC, can incrementally dock a peptide in the binding site and does not require any template (*see* Fig. 2 and Supplemental Material).[9,10,21] Each approach has different strengths and limitations, and can therefore suit various user needs, depending on the task at hand. For instance, its speed makes APE-Gen better suited for large-scale modeling and structure-based virtual screening. On the other hand, since it does not rely on templates, DINC's predictions can be more general and account for unusual binding modes, thus making it more suited for geometry prediction.[9,22] Both APE-Gen and DINC have been validated in previous publications.[8,10,23] In this paper, we present a unified environment that facilitates the utilization of APE-Gen, DINC and other tools, for various research applications.

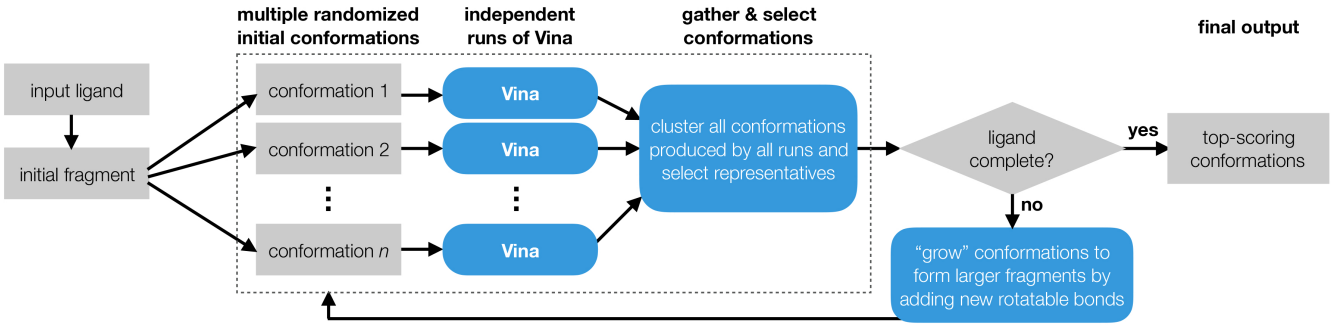


Figure 2. Workflow of DINC's parallel and incremental meta-docking approach. DINC starts by selecting a small fragment of the input ligand, with only k flexible bonds. Multiple conformations are created by randomly sampling different values for the dihedral angles of this fragment. These n conformations are then used as input for multiple independent runs of a docking tool, in this example, Vina,[20] which are executed in parallel by different threads. From all the binding modes produced by these parallel runs, the n "best" ones are selected for expansion: they are "grown" by adding several atoms and bonds from the input ligand. These larger fragments are then docked independently, in parallel, while keeping the number of flexible bonds equal to k. This process is repeated until the entire input ligand has been incrementally reconstructed, and is docked in the receptor's binding site.

HLA-Arena: Structural modeling and analysis of pHLA complexes

Using Jupyter Notebook and Docker, we have created a customizable environment, called HLA-Arena, that enables researchers to easily model any class I pHLA complex of interest and perform varied structural analyses (*see* Fig. 3). HLA-Arena includes different workflows, defined as separate notebooks, that consist of the following main stages:

- **Input processing:** Available structures of HLA receptors are obtained from the PDB[26] to be used as such or as templates. Unavailable HLA structures are modeled with Modeller,[11] using a HLA sequence and the structure of a similar HLA receptor as template, if these are provided by the user. Alternatively, users can just provide an allele name (e.g., HLA-A*24:02); HLA-Arena will then fetch the proper sequence from IMGT/HLA,[7] and a reasonable template (based on the HLA supertype[27] classification) from the PDB. In addition, binding affinity of peptides can be estimated with MHCflurry[12] to select the most relevant ones.
  Minimal example: `HLA_allele = arena.model_hla('HLA-A*24:02')`

- **Peptide docking:** Structures of pHLA complexes are modeled with APE-Gen and/or DINC, which only requires the sequence of the target peptide(s) and the HLA structure(s) obtained previously. Modeled structures can also be minimized with a force field, using OpenMM.[24]
  Minimal example: `structure = arena.dock('QFKDNVILL', HLA_allele)`

- **Data analysis:** A variety of post-processing options for data analysis can be incorporated in a workflow. These include binding mode rescoring or peptide ranking with DINC, and structure visualization with NGL Viewer,[13] among others.
  Minimal example: `arena.visualize(structure)`

For a smooth user experience, all computational tools involved in HLA-Arena are packaged within a Docker image (*see* Supplemental Material for installation details), therefore eliminating the burden of managing software dependencies. Another advantage of Docker containerization is to make HLA-Arena platform-agnostic. As a result, it can be deployed on a desktop computer or a high-performance computing

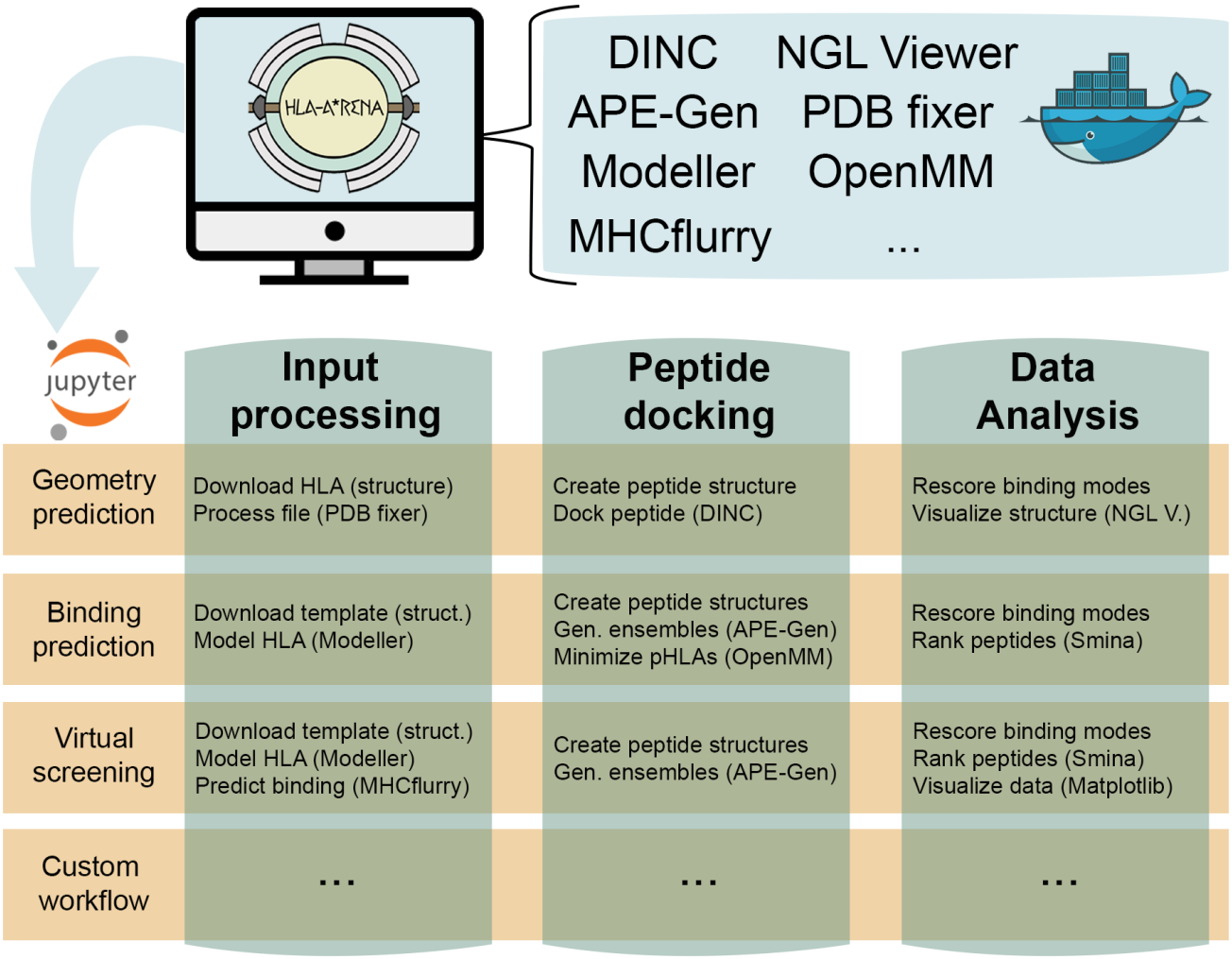


Figure 3. HLA-Arena leverages Docker and Jupyter Notebook, offering a customizable environment to build and execute various workflows for the structural modeling and analysis of pHLA complexes. Three proposed workflows are depicted here: geometry prediction of pHLA binding modes, structure-based prediction of binding energy, and virtual screening of tumor-derived peptides. In the geometry prediction workflow, after obtaining the structure of a HLA receptor, a peptide of interest is docked in its binding site by DINC, and all generated binding modes are scored with several scoring functions. In the binding prediction workflow, after modeling a given HLA structure, ensembles of binding modes are generated with APE-Gen (and optionally minimized with OpenMM)[24] for various peptides, and these binding modes are scored to rank the peptides with Smina.[25] In the virtual screening workflow, after filtering peptides with MHCflurry,[12] ensembles of binding modes are generated with APE-Gen for the selected peptides, and the top-scoring binding modes are used to rank these peptides with Smina, in terms of binding affinity to a (set of) HLA receptor(s). Note that these workflows can be modified, and that new ones can be created by users. In each application, different types of data analysis can be used to guide the selection of the best pHLAs before experimental validation.

cluster, across different operating systems. Users can customize available workflows by adding modeling or analysis steps. We plan to continuously expand the capabilities of HLA-Arena by providing support for additional tools.[28–30]

Results

We now present the results we obtained when carrying out three different workflows that exemplify the diversity of applications offered by HLA-Arena. Each workflow leverages the functionalities of several tools in a coherent scenario.

Geometry prediction of pHLA binding modes

HLA-Arena can be used to predict conformations of peptides bound to HLA receptors even for peptides presenting unusual binding modes.[10] To illustrate this, using the geometry prediction workflow based on DINC (*see* Fig. 3), we try and reproduce the crystal structures of two such peptides.

First, we conduct a self-docking experiment with a crystal structure (with PDB code 1E27) involving HLA-B*51:01 and a 9-mer peptide derived from HIV-1. It has been suggested that the fifth residue acts as a secondary anchor for this peptide, leading to structural rearrangement of its central and amino-terminal residues.[31] Our experiment evaluates the capability of DINC to reproduce the bound geometry of this peptide, without considering receptor flexibility. To evaluate performance and reproducibility, we carry out this experiment with either 8 or 32 threads (for the parallel process in DINC), running five replicates in each case. Default values are used for other DINC parameters.[23] Results (*see* Fig. S1.A) show that, in every single run, HLA-Arena sampled a near-native peptide conformation, i.e., a conformation with an all-heavy-atom RMSD (root mean square deviation) to the crystal structure less than 2.5 Å.

Geometry prediction involves two issues that are especially challenging with peptides.[5] The first relates to *sampling*, i.e., how to explore the full flexibility of a large ligand. The second relates to *scoring*, i.e., how to identify the "best" ligand conformation in a pool of diverse binding modes. HLA-Arena relies on the incremental process of DINC to overcome the sampling issue. It also includes a filtering step to remove peptide conformations with reverse orientation in the binding cleft. To address the scoring issue, HLA-Arena makes use of multiple scoring functions. For instance, in this self-docking experiment, conformations were ranked with the scoring functions of AutoDock4,[32] Vina[20] and Vinardo.[33] All three scoring functions were able to identify near-native conformations. However, only in the case of AutoDock4 (*see* Fig. 4.A), the top five ranking conformations in one of the replicates included the overall lowest-RMSD conformation, i.e., the conformation with the lowest-RMSD to the crystal structure among all sampled conformations (*see* Fig. S1.A).

Second, we try and reproduce a crystal structure (with PDB code 2GTW) involving HLA-A*02:01 and a 9-mer peptide derived from the MART-1/Melan-A protein.[34] This peptide has an A27L substitution in comparison to the MART-1 peptide targeted by numerous clinical studies.[35,36] This substitution leads to an alternative arrangement of primary anchor residues, resulting in an unusual binding mode.[10,29,34] Again, we run five replicates of the geometry prediction workflow, using either 8 or 32 threads. For the prediction task to be closer to a real-case scenario, we perform a cross-docking experiment, accounting for receptor flexibility. It makes this task much harder, from both the sampling and scoring perspectives.[37,38] In spite of that, HLA-Arena sampled near-native conformations, although it performed better when using 32 threads (*see* Fig. S1.B and S2). In terms of scoring, only AutoDock4 and Vinardo were able to recover near-native conformations (*see* Fig. 4.B). Note that HLA-Arena also allows visualizing the three-dimensional structure of the top-ranking binding mode (*see* Fig. 4.C and D).

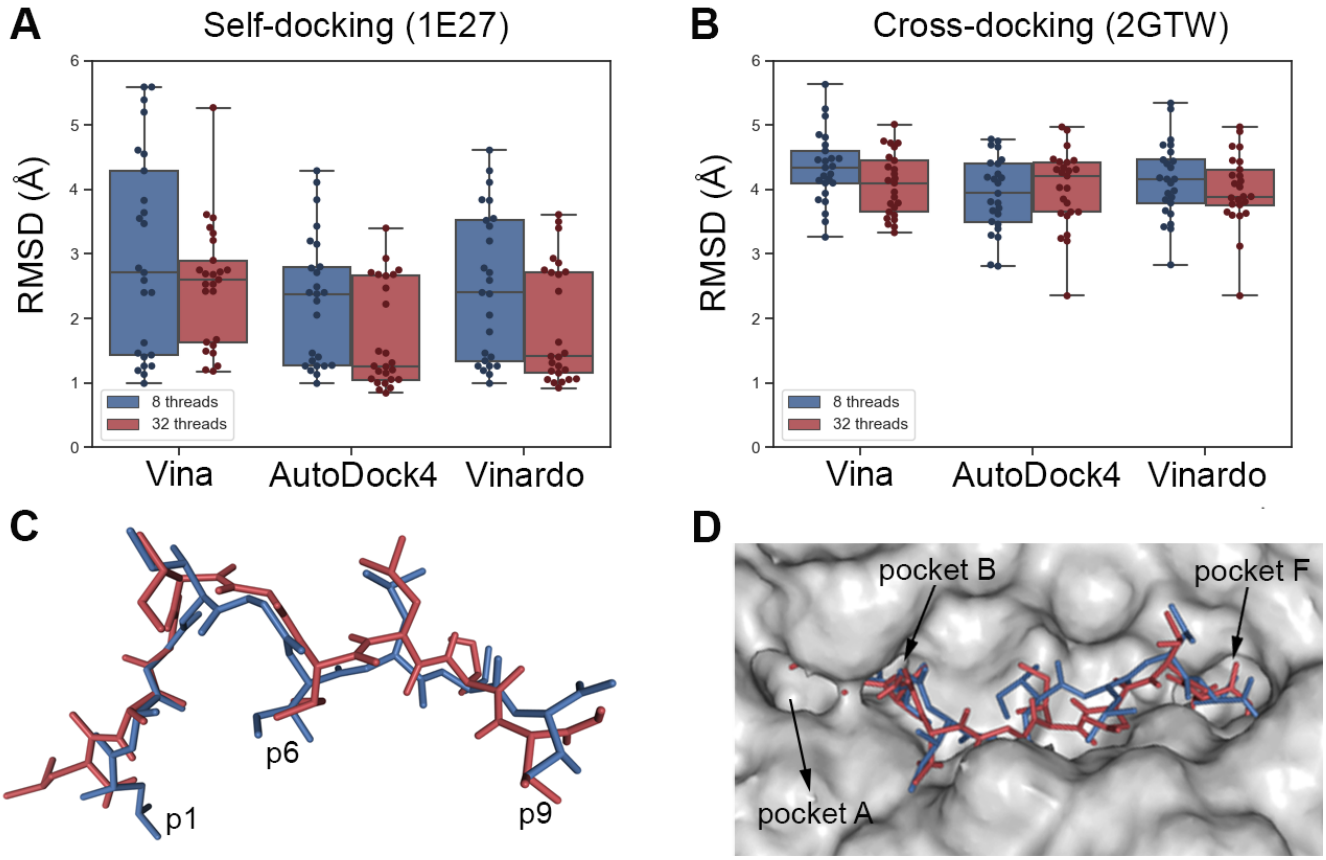


Figure 4. Geometry prediction of pHLA binding modes. **A**. Three scoring functions are used to select the top 5 ranking conformations produced by five replicates of a self-docking experiment aimed at predicting the binding mode of a 9-mer peptide (under PDB code 1E27), using 8 or 32 threads for DINC. Each box plot aggregates results of the five replicates. Each dot corresponds to a conformation, plotted according to its all-heavy-atom RMSD (root mean square deviation, in Å) to the reference crystal structure. **B**. Results of a cross-docking experiment aimed at predicting the binding mode of a 9-mer peptide (under PDB code 2GTW), obtained with the same methodology. **C**. Side view of the best binding mode (in red), identified by AutoDock4 and Vinardo, and aligned with the crystal structure (in blue) of this peptide (under PDB code 2GTW). Only heavy atoms are depicted, using a *sticks* representation. Note that this sampled conformation has an all-heavy-atom RMSD of 2.35 Å, and does not perfectly reproduce the side-chain arrangement of the first residue. A better conformation, with an all-heavy-atom RMSD of 2.15 Å, was sampled by HLA-Arena (*see* Fig. S2), but was not among the top ranking conformations. **D**. Top view of the HLA binding site (depicted by a grey surface) with peptide conformations shown in C within it (as sticks). This peptide uses its first amino acid as primary anchor (i.e., residue p1 is anchored in pocket B), which is quite unusual for HLA-A*02:01 binders. Images in C and D were generated with HLA-Arena using the embedded NGL Viewer.[13] Both images were edited to add labels.

Structure-based prediction of binding energy

To demonstrate another application of HLA-Arena, we use the binding prediction workflow (*see* Fig. 3) to predict binding to HLA-A*02:01, for a small dataset of selected peptides (*see* Table S1). This dataset includes five experimentally-identified non-binders, as well as 11 binders whose experimental binding affinities are available in the Immune Epitope Database (IEDB),[39] and whose crystal structures in complex with HLA-A*02:01 are available in the PDB. For each peptide, we generate an ensemble of bound conformations with APE-Gen. Each peptide's binding energy is then estimated as the median score within the conformation ensemble, for each scoring function (namely AutoDock4, Vina and Vinardo). Correlations between these predicted binding energies and experimentally-determined binding affinities are then determined (*see* Fig. 5).

In addition to the default local optimization performed by APE-Gen, HLA-Arena provides the option of minimizing the resulting complexes with OpenMM.[24] To evaluate the impact of this procedure, we recalculate binding energies and correlations after running this energy minimization for all conformations in each ensemble. Our results show a consistent increase of the predicted binding energies, for all scoring functions (*see* Fig. 5). This might reflect the differences in binding energy estimation that exist between these empirical or semi-empirical scoring functions[40] and the force field used by OpenMM (i.e., amber99sbildn).[24] Despite increasing binding energies, the OpenMM minimization has a positive impact on overall correlations.

Interestingly, the best correlation with experimental binding affinities is obtained when using Vina. This result is in agreement with previous studies evaluating Vina's performance for virtual screening of drug-like ligands.[40,41] Note that contrary to the geometry prediction workflow, in which a scoring function is only used to rank different conformations of a given peptide, here, the scoring function also has to rank different peptides. Although the same function can be used for both purposes,[5] it is possible that better results are obtained when using functions optimized for each task.

For the HLA-A*02:01–binders in our dataset, we can compute RMSDs between their associated crystal structures and conformations generated by APE-Gen. This allows verifying that APE-Gen ensembles include near-

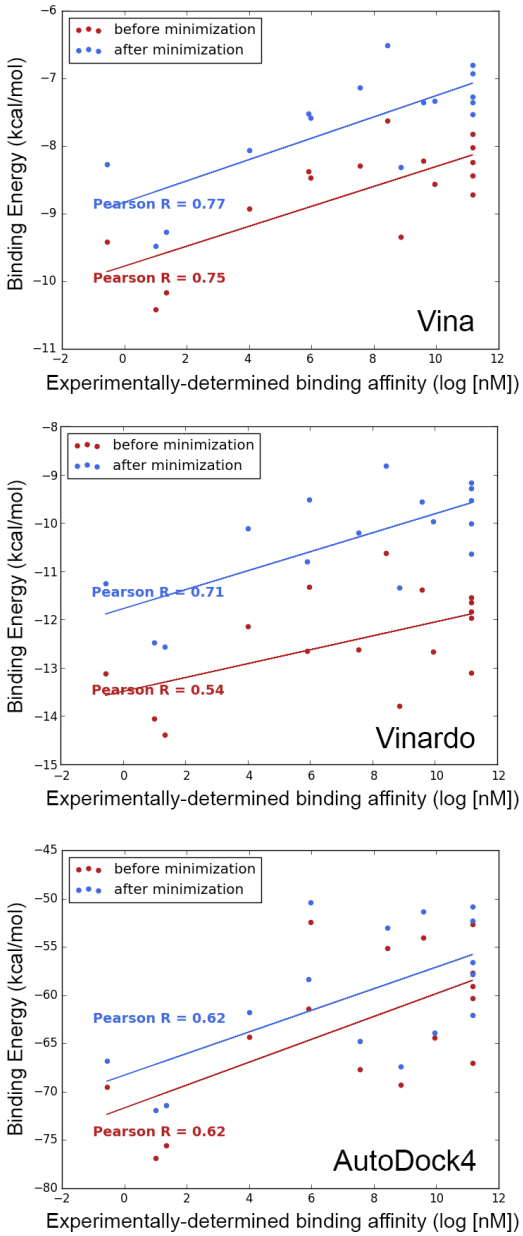


Figure 5. HLA-A*02:01–binding predicted for a small set of peptides. Each plot illustrates the correlation between experimentally-obtained binding affinities (extracted from the Immune Epitope Database) and structure-based binding energies, as predicted by a given scoring function (namely AutoDock4, Vina or Vinardo). Structures were generated with APE-Gen, with or without minimization with OpenMM (see text for details). Correlation coefficients are also reported (as Pearson's R). Each point corresponds to a peptide in Table S1.

native conformations (*see* Fig. S3), and evaluating the impact of the OpenMM minimization on these conformations. This also allows comparing the use of an ensemble of conformations, to predict binding energies, with the use of a single conformation from this ensemble (e.g., the conformation with the lowest RMSD to the corresponding crystal structure). Our results with Vina's scoring function suggest that better correlations are obtained with ensembles of conformations (*see* Fig. S4).

Virtual screening of tumor-derived peptides

HLA-Arena allows researchers to perform, for the first time, a large-scale structure-based virtual screening of HLA-binding peptides. In addition, by combining sequence-based and structure-based methods, HLA-Arena represents a fresh alternative for the identification of tumor-derived peptide-targets, considering patient-specific HLAs. To demonstrate this application, we use the virtual screening workflow (*see* Fig. 3) to predict which peptides are the strongest binders to the class I HLA receptors of a fictitious cancer patient.

We consider six alleles: HLA-A*24:02, HLA-A*26:01, HLA-B*15:01, HLA-B*35:01, HLA-C*04:01 and HLA-C*05:01. We build a peptide dataset by selecting 500 known binders and 1,000 decoys for each allele, for a total of 9,000 peptides. Sequences of known binders are obtained from SysteMHC Atlas,[42] where they have been derived from immunopeptidomics studies. Sequences of decoys are obtained from the training set of NetMHCpan.[43]

First, the whole dataset of peptides is screened for HLA binding with MHCflurry,[12] using an affinity threshold specified by the user. This allows quickly selecting the most likely binders for each HLA receptor, before proceeding with the more computationally expensive steps. In this example, a threshold of 500 nM selects 2,604 peptides. Then, we proceed with the structural modeling of the full pHLA complex for all selected peptides. Finally, peptides are ranked based on binding energies derived from the modeled structures. The entire pipeline takes ≈86 hours on a desktop computer, or ≈5 hours on a high-performance cluster (*see* Supplemental Material).

The threshold used in MHCflurry directly impacts the sensitivity/specificity of the overall prediction. Recent surveys indicate that commonly used thresholds for sequence-based HLA-binding predictors (e.g., 500 nM) can yield a sensitivity as low as 40%,[44] with great variation in accuracy between HLA alleles.[45] On our dataset, a 500 nM threshold produces several false positive predictions (*see* the blue dots in Fig. 6.A) and false negative predictions (data not shown). While trying to address this issue, we have observed that our structure-based analysis can usually eliminate at least half of false positive predictions, and recover significant numbers of false negative predictions, although results vary depending on the studied HLA allele (data not shown).

As our workflow allows varying the MHCflurry threshold, we repeat the aforementioned virtual screening experiment with a 50,000 nM value. This leads to all 9,000 peptides being selected for modeling and ranking. The observed enrichment of true binders among the top ranking peptides (i.e., the red dots at the bottom of the distributions in Fig. 6.B) further corroborates our claim that structural information is useful when screening HLA-binders.

In these examples, we performed only one sampling round in APE-Gen for each complex, and only the top-scored conformation was used as input for ranking. Better results could be obtained by (i) executing more sampling rounds in APE-Gen, (ii) performing the OpenMM minimization, or (iii) using the whole APE-Gen ensemble. More importantly, accurate scoring remains an open challenge. Therefore, structure-based predictions cannot yet outperform sequence-based methods, but can be combined with them to provide additional information when selecting peptides for experimental validation.

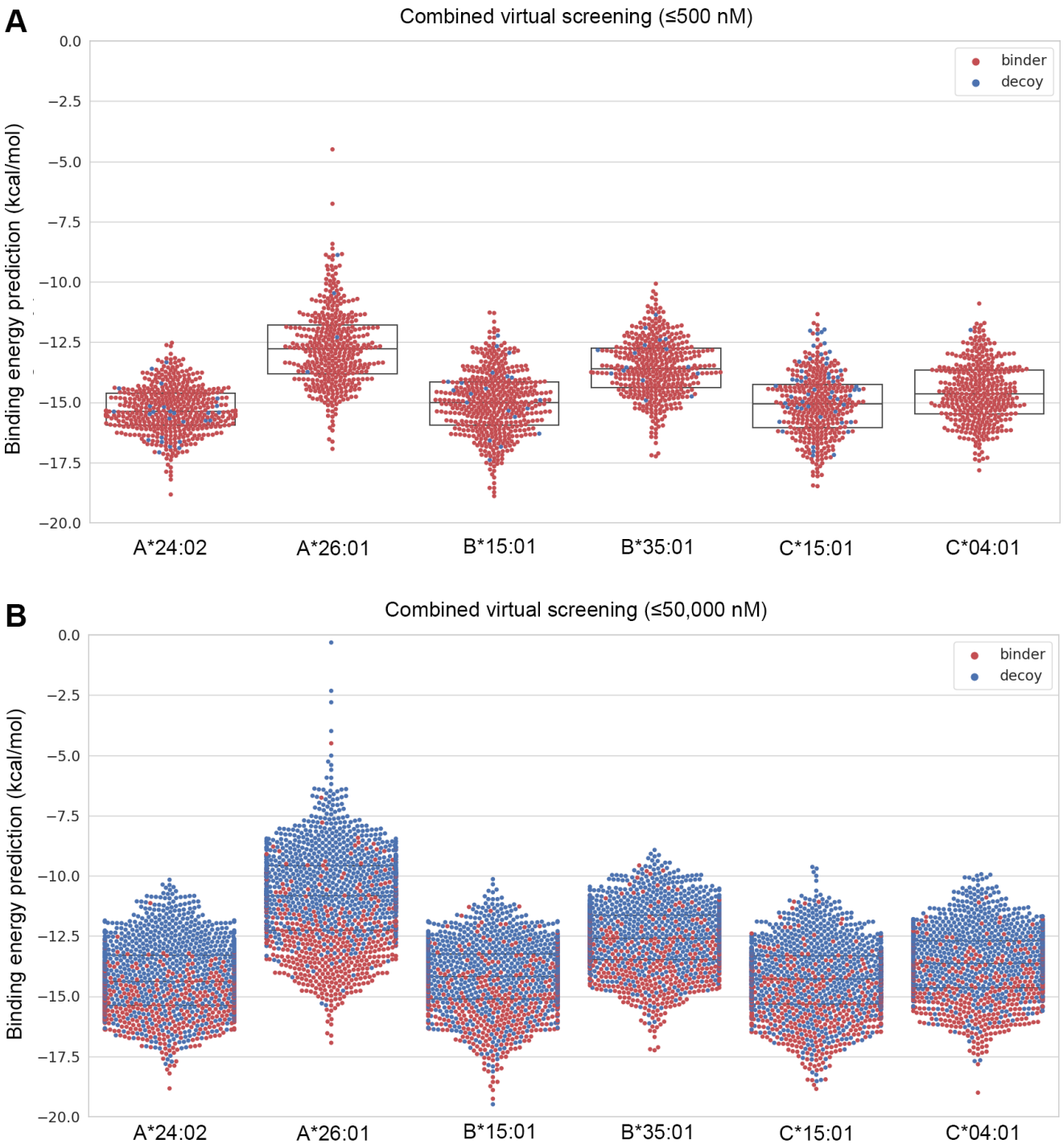


Figure 6. Structure-based virtual screening for high-affinity HLA binders. The HLA-Arena virtual screening workflow was used to predict peptide binders for 6 HLA receptors of interest. For this exercise, a dataset of 9,000 peptides was created, using 500 known binders (red dots) and 1000 decoys (blue dots) for each HLA. **A**. Results of a combined virtual screening (i.e., MHCflurry + APE-Gen) with a 500 nM threshold for MHCflurry. **B**. Results of the same virtual screening using a 50,000 nM threshold for MHCflurry. In both plots, each dot corresponds to the top-scoring conformation of a modeled pHLA complex, selected from the ensemble of conformations produced by APE-Gen. For each HLA (on the x-axis), complexes with the lowest binding energies (on the y-axis) would be predicted as the best candidates for further analysis or experimental validation.

Conclusion

HLA-Arena provides researchers with a customizable environment to create and execute sophisticated workflows for the structural modeling and analysis of pHLA complexes. Its intuitive interface relies on Jupyter Notebook and Docker to dramatically reduce the burden of software dependencies and the need for advanced programming skills, making its resources accessible to a wide audience. Available workflows combine commonly-used software for protein modeling and analysis, with tools that we developed to address challenges specific to pHLA complexes. We believe that HLA-Arena could become a stepping-stone for a broad collaborative effort to study pHLA complexes.

In this paper, we have presented three workflows to showcase the capabilities of HLA-Arena. First, HLA-Arena enables the geometry prediction of pHLA structures, even for peptides with unusual binding modes, by using template-free molecular docking. Second, HLA-Arena allows predicting binding energies for potential HLA binders, by quickly producing ensembles of bound conformations for these peptides and rescoring all the results. Third, HLA-Arena enables a more accurate virtual screening of HLA binders, by combining sequence-based and structure-based approaches.

These workflows can be modified to allow for additional analysis of the modeled pHLA complexes, for example to perform molecular dynamics with OpenMM[46,47] or cross-reactivity assessment.[2,48,49] Thanks to high-performance computing and efficient sampling, molecular dynamics could play a bigger role in providing accurate estimates of pHLA binding affinity and complex stability.[50,51]

HLA-Arena can be integrated in computational pipelines for basic cancer research, or help inform physicians in pre-clinical settings. It can be used to perform the large-scale modeling and selection of tumor-associated peptides, the computer-aided design of altered peptide ligands, and the study of T-cell cross-reactivity.[2,8] In addition to HLA binding prediction, immunotherapy applications require identifying peptides that are uniquely displayed by cancer cells. This important task will be addressed in future updates of HLA-Arena.

It is important to note that HLA-Arena provides efficient solutions to sampling challenges associated with pHLA modeling,[8,23] and facilitates the integration of these solutions with other tools for structural analysis. However, the accuracy of structure-based peptide ranking is limited by existing scoring functions. As they improve, new scoring functions will be incorporated in HLA-Arena to replace current ones or be combined into consensus methods.[52,53] In time, we expect that structure-based analyses will become essential to peptide-target prediction for neoantigen discovery, vaccine development, and cancer immunotherapy, specially for patients with less prevalent HLA alleles.

Data Availability Statement

HLA-Arena is made available through Docker Hub, under *kavrakilab/hla-arena* (*see* Supplemental Material for installation details). The HLA-Arena Docker image also contains data related to the experiments described here, which can be reproduced as demo workflows. Additional information and Documentation can also be found on GitHub, at https://github.com/KavrakiLab/hla-arena.

Support

This work was funded in part by NIH (grant number 1R21CA209941-01), through the Informatics Technology for Cancer Research (ITCR) initiative of the National Cancer Institute (NCI). This work was also partially supported by the Cancer Prevention & Research Institute of Texas (CPRIT), through award number RP170508, and through a fellowship from the Gulf Cost Consortia on the Computational Cancer Biology Training Program (grant number RP170593). Finally, this work was also partially supported by a

training fellowship from the National Library of Medicine Training Program in Biomedical Informatics (grant number T15LM007093), and by Rice University funds.

Acknowledgements

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation (NSF) grant number ACI-1548562. More specifically, it involved the Stampede cluster, at the Texas Advanced Computing Center (TACC), through allocation MCB180187. This work was also supported by the Big-Data Private-Cloud Research Cyberinfrastructure MRI-award, funded by NSF under grant CNS-1338099, and by Rice University. We also thank Romanos Fasoulis for his contributions in testing some of the components of HLA-Arena. Finally, the computer image on Fig. 3 was modified from an icon designed by *ultimatearm* from Flaticon.

Author contributions

Conception and design: Dinler Antunes, Jayvee Abella, Sarah Hall-Swan, Didier Devaurs, Mark Moll, Gregory Lizée, Lydia Kavraki
Financial support: Gregory Lizée, Lydia Kavraki
Administrative support: Gregory Lizée, Lydia Kavraki
Collection and assembly of data: Jayvee Abella, Dinler Antunes, Sarah Hall-Swan, Anja Conev
Data analysis and interpretation: Dinler Antunes, Jayvee Abella, Didier Devaurs
Manuscript writing: Dinler Antunes, Didier Devaurs
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

Authors' disclosures of potential conflicts of interest

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript.

Lydia E. Kavraki

Consulting or advisory role: Allgens Biomedical Research Co., Ltd; Biomedical and Informatics Consultants, Inc. (I)
Patents, royalties, other intellectual property: Rice University (I)
Expert testimony: Dialysis Clinic, Inc.; Maynard, Cooper & Gale (I)
Travel, accommodations, expenses: Beijing Advanced Medical Technologies, Ltd. Inc. (I)

No other potential conflicts of interest were reported.

References

1. Lizée, G., Overwijk, W. W., Radvanyi, L. *et al.* Harnessing the power of the immune system to target cancer. *Annu Rev Med* **64**, 71–90 (2013).
2. Antunes, D. A., Rigo, M. M., Freitas, M. V. *et al.* Interpreting T-cell cross-reactivity through structure: Implications for TCR-based cancer immunotherapy. *Front Immunol* **8**, 1210 (2017).

3. Degauque, N., Brouard, S. & Soulillou, J. P. Cross-reactivity of TCR repertoire: Current concepts, challenges, and implication for allotransplantation. *Front Immunol* **7**, 89 (2016).

4. Adams, J. J., Narayanan, S., Birnbaum, M. E. *et al.* Structural interplay between germline interactions and adaptive recognition determines the bandwidth of TCR-peptide-MHC cross-reactivity. *Nat Immunol* **17**, 87–94 (2016).

5. Antunes, D. A., Abella, J. R., Devaurs, D. *et al.* Structure-based methods for binding mode and binding affinity prediction for peptide-MHC complexes. *Curr Top Med Chem* **18**, 2239–2255 (2018).

6. Vandiedonck, C. & Knight, J. C. The human Major Histocompatibility Complex as a paradigm in genomics research. *Brief Funct Genomic Proteomic* **8**, 379–394 (2009).

7. Robinson, J., Halliwell, J. A., Hayhurst, J. D. *et al.* The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res* **43**, D423–431 (2015).

8. Abella, J. R., Antunes, D. A., Clementi, C. *et al.* APE-Gen: A fast method for generating ensembles of bound peptide-MHC conformations. *Molecules* **24**, 881 (2019).

9. Antunes, D. A., Moll, M., Devaurs, D. *et al.* DINC 2.0: A new protein-peptide docking webserver using an incremental approach. *Cancer Res* **77**, e55–e57 (2017).

10. Antunes, D. A., Devaurs, D., Moll, M. *et al.* General prediction of peptide-MHC binding modes using incremental docking: A proof of concept. *Sci Rep* **8**, 4327 (2018).

11. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* **86**, 1–2 (2016).

12. O'Donnell, T. J., Rubinsteyn, A., Bonsack, M. *et al.* MHCflurry: Open-source class I MHC binding affinity prediction. *Cell Syst* **7**, 129–132 (2018).

13. Rose, A. S., Bradley, A. R., Valasatava, Y. *et al.* NGL viewer: Web-based molecular graphics for large complexes. *Bioinformatics* **34**, 3755–3758 (2018).

14. Dos Santos Francisco, R., Buhler, S., Nunes, J. M. *et al.* HLA supertype variation across populations: New insights into the role of natural selection in the evolution of HLA-A and HLA-B polymorphisms. *Immunogenetics* **67**, 651–663 (2015).

15. Wieczorek, M., Abualrous, E. T., Sticht, J. *et al.* Major Histocompatibility Complex (MHC) class I and MHC class II proteins: Conformational plasticity in antigen presentation. *Front Immunol* **8**, 292 (2017).

16. Fodor, J., Riley, B. T., Borg, N. A. & Buckle, A. M. Previously hidden dynamics at the TCR-peptide-MHC interface revealed. *J Immunol* **200**, 4134–4145 (2018).

17. Bordner, A. J. & Abagyan, R. Ab initio prediction of peptide-MHC binding geometry for diverse class I MHC allotypes. *Proteins* **63**, 512–526 (2006).

18. Khan, J. M. & Ranganathan, S. pDOCK: A new technique for rapid and accurate docking of peptide ligands to Major Histocompatibility Complexes. *Immunome Res* **6**, S2 (2010).

19. Chys, P. & Chacón, P. Random coordinate descent with spinor-matrices and geometric filters for efficient loop closure. *J Chem Theory Comput* **9**, 1821–1829 (2013).

20. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455–461 (2010).

21. Dhanik, A., McMurray, J. S. & Kavraki, L. E. DINC: A new AutoDock-based protocol for docking large ligands. *BMC Struct Biol* **13**, S11 (2013).

22. Dhanik, A., McMurray, J. S. & Kavraki, L. E. Binding modes of peptidomimetics designed to inhibit STAT3. *PLoS One* **7**, e51603 (2012).

23. Devaurs, D., Antunes, D. A., Hall-Swan, S. *et al.* Using parallelized incremental meta-docking can solve the conformational sampling issue when docking large ligands to proteins. *BMC Mol Cell Biol* **20**, 42 (2019).

24. Eastman, P., Swails, J., Chodera, J. D. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* **13**, e1005659 (2017).

25. Koes, D. R., Baumgartner, M. P. & Camacho, C. J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J Chem Inf Model* **53**, 1893–1904 (2013).

26. Berman, H. M., Westbrook, J., Feng, Z. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).

27. Sidney, J., Peters, B., Frahm, N. *et al.* HLA class I supertypes: A revised and updated classification. *BMC Immunol* **9**, 1 (2008).

28. Liu, T., Pan, X., Chao, L. *et al.* Subangstrom accuracy in pHLA-I modeling by Rosetta FlexPepDock refinement protocol. *J Chem Inf Model* **54**, 2233–2242 (2014).

29. Rigo, M. M., Antunes, D. A., Vaz de Freitas, M. *et al.* DockTope: A web-based tool for automated pMHC-I modelling. *Sci Rep* **5**, 18413 (2015).

30. Kyeong, H. H., Choi, Y. & Kim, H. S. GradDock: Rapid simulation and tailored ranking functions for peptide-MHC class I docking. *Bioinformatics* **34**, 469–476 (2018).

31. Maenaka, K. *et al.* Nonstandard peptide binding revealed by crystal structures of HLA-B*5101 complexed with HIV immunodominant epitopes. *J. Immunol.* **165**, 3260–3267 (2000).

32. Morris, G. M., Huey, R., Lindstrom, W. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **30**, 2785–2791 (2009).

33. Quiroga, R. & Villarreal, M. A. Vinardo: A scoring function based on Autodock Vina improves scoring, docking, and virtual screening. *PLoS ONE* **11**, e0155183 (2016).

34. Borbulevych, O. Y., Insaidoo, F. K., Baxter, T. K. *et al.* Structures of MART-126/27-35 Peptide/HLA-A2 complexes reveal a remarkable disconnect between antigen structural homology and T cell recognition. *J Mol Biol* **372**, 1123–1136 (2007).

35. Kawakami, Y., Eliyahu, S., Sakaguchi, K. *et al.* Identification of the immunodominant peptides of the MART-1 human melanoma antigen recognized by the majority of HLA-A2-restricted tumor infiltrating lymphocytes. *J Exp Med* **180**, 347–352 (1994).

36. Rivoltini, L., Kawakami, Y., Sakaguchi, K. *et al.* Induction of tumor-reactive CTL from peripheral blood and tumor-infiltrating lymphocytes of melanoma patients by in vitro stimulation with an immunodominant peptide of the human melanoma antigen MART-1. *J Immunol* **154**, 2257–2265 (1995).

37. Antunes, D. A., Devaurs, D. & Kavraki, L. E. Understanding the challenges of protein flexibility in drug design. *Expert Opin Drug Discov* **10**, 1301–1313 (2015).

38. Wang, Z., Sun, H., Yao, X. *et al.* Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: The prediction accuracy of sampling power and scoring power. *Phys Chem Chem Phys* **18**, 12964–12975 (2016).

39. Vita, R., Overton, J. A., Greenbaum, J. A. *et al.* The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* **43**, D405–412 (2015).

40. Nguyen, N. T. *et al.* Autodock Vina Adopts More Accurate Binding Poses but Autodock4 Forms Better Binding Affinity. *J Chem Inf Model* **60**, 204–211 (2020).

41. v, M. W., Ayeni, C., Breuer, S. & Torbett, B. E. Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina. *PLoS ONE* **5**, e11955 (2010).

42. Shao, W., Pedrioli, P. G., Wolski, W. *et al.* The SysteMHC Atlas project. *Nucleic Acids Res* **46**, D1237–D1247 (2018).

43. Jurtz, V. *et al.* NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J. Immunol.* **199**, 3360–3368 (2017).

44. Bonsack, M. *et al.* Performance Evaluation of MHC Class-I Binding Prediction Tools Based on an Experimentally Validated MHC-Peptide Binding Data Set. *Cancer Immunol Res* **7**, 719–736 (2019).

45. Zhao, W. & Sher, X. Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes. *PLoS Comput. Biol.* **14**, e1006457 (2018).

46. Wieczorek, M., Sticht, J., Stolzenberg, S. *et al.* MHC class II complexes sample intermediate states along the peptide exchange pathway. *Nat Commun* **7**, 13224 (2016).

47. Pang, Y. P., Elsbernd, L. R., Block, M. S. & Markovic, S. N. Peptide-binding groove contraction linked to the lack of T cell response: Using complex structure and energy to identify neoantigens. *Immunohorizons* **2**, 216–225 (2018).

48. Mendes, M. F., Antunes, D. A., Rigo, M. M. *et al.* Improved structural method for T-cell cross-reactivity prediction. *Mol Immunol* **67**, 303–310 (2015).

49. Dhanik, A., Kirshner, J. R., MacDonald, D. *et al.* In-silico discovery of cancer-specific peptide-HLA complexes for targeted therapy. *BMC Bioinformatics* **17**, 286 (2016).

50. Wan, S., Knapp, B., Wright, D. W. *et al.* Rapid, precise, and reproducible prediction of peptide-MHC binding affinities from molecular dynamics that correlate well with experiment. *J Chem Theory Comput* **11**, 3346–3356 (2015).

51. Ochoa, R., Laio, A. & Cossio, P. Predicting the affinity of peptides to Major Histocompatibility Complex class II by scoring molecular dynamics simulations. *J Chem Inf Model* **59**, 3464–3473 (2019).

52. Kukol, A. Consensus virtual screening approaches to predict protein ligands. *Eur J Med Chem* **46**, 4661–4664 (2011).

53. Palacio-Rodríguez, K., Lans, I., Cavasotto, C. N. & Cossio, P. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. *Sci Rep* **9**, 5142 (2019).

Supplemental Material

APE-Gen: Fast generation of pHLA binding mode ensembles

We recently released a new tool, the anchored peptide-HLA ensemble generator (APE-Gen), that produces an ensemble of binding modes for a pHLA complex, starting from the sequences of a peptide and HLA receptor.[8] APE-Gen involves an iterative process repeating the three following steps. First, the ends of the peptide's backbone are anchored within known pockets in the HLA's binding site, using available backbone termini templates. Second, the peptide backbone is completed by applying the random coordinate descent (RCD) loop-modeling tool,[19] which efficiently yields several valid backbone conformations. Third, side chains are added to the backbone conformations, and local optimization is performed with Smina[25] to fix steric clashes. This step considers full peptide flexibility and binding-site side-chains flexibility, producing a set of full-atom peptide conformations within the HLA's binding site. After each such round of sampling, the highest-quality conformation (according to the internal scoring function, currently Vinardo) can be used as template for the next round.

By generating a diverse ensemble of pHLA binding modes, APE-Gen implicitly accounts for the natural flexibility of peptides within the binding site. We have shown that APE-Gen could reproduce the entire set of non-redundant classical class I pHLA structures available in the protein data bank (PDB),[26] i.e., 535 complexes at the time of the study.[8] In that case, we used a single round of sampling per complex. The average root mean square deviation (RMSD) between modeled peptides and their corresponding crystal structure (considering all heavy atoms) was of only 2.02 Å, which is considered an accurate reproduction. Even better results can be obtained when performing optimization and/or additional rounds of sampling, specially for longer peptides.[8]

APE-Gen is very computationally-efficient, producing dozens of binding modes in a few minutes on a standard desktop computer. It can be run for several peptides and a given HLA receptor, therefore producing valuable information for peptide ranking and binding affinity prediction, and enabling structure-based virtual screening of HLA-binding peptides. We have also shown the potential benefits of APE-Gen when studying T-cell cross-reactivity.[8]

DINC: Incremental docking of pHLA complexes

In previous work, we presented a molecular docking approach, called DINC (which stands for *docking incrementally*), specifically developed for large ligands, including peptides.[21] The underlying idea is to incrementally dock larger and larger fragments of a ligand, instead of trying to dock it all at once. Note that this incremental docking process focuses on ligand flexibility, although selected receptor side-chains can also be sampled. This process is parallelized to allow for broader sampling, by having several runs of docking performed independently at each step, and grouping their results together. DINC is also a meta-docking method, in the sense that it relies on existing molecular docking tools, such as AutoDock4,[32] Vina[20] and Smina[25] to perform the docking of the fragments at each step. As a consequence, fragment sampling and scoring can be performed by different tools.

The latest version of our software, called DINC 2.0, has been made available as a web server.[9] We have recently shown that it performs a more exhaustive sampling than other docking approaches.[23] In that study, DINC was benchmarked using five public datasets including large ligands; it reproduced many crystal structures on which other docking tools had failed.[23] For example, it has been used to study the inhibition of the Src homology 2 domain of STAT3 by peptidomimetics.[22] We have also shown that DINC could reproduce a diverse set of pHLA structures encompassing ten HLA alleles and peptides with diverse binding modes; it achieved an average all-heavy-atom RMSD of 1.92 Å.[10] Note that DINC is not limited to common class I HLA receptors, contrary to many related tools.[5] It can be applied to complexes involving

synthetic ligands, rare and non-classical class I HLAs, and potentially to class II HLA receptors.[9] An updated version of DINC is made available through Docker Hub (*docker pull kavrakilab/dinc-bin*).

HLA-Arena performance for virtual screening

HLA-Arena provides the most efficient workflow available for structure-based virtual screening of HLA-binders. For the experiment we report in the Results section, the breakdown of computing time is as follows: MHCflurry[12] needs about 15 seconds to screen the entire dataset of 9,000 peptides. The homology modeling step takes about 3 minutes for each HLA allele, and can be skipped for HLAs with available crystal structures. The APE-Gen step takes about 2 minutes per pHLA complex, on a desktop computer with 6-8 threads. The (optional) rescoring takes about 2 seconds per complex, using an HLA-Arena function that relies on Smina.[25] Therefore, running the entire workflow on a desktop computer takes about 86 hours with an MHCflurry threshold at 500 nM, and about 300 hours with an MHCflurry threshold at 50,000 nM. This running time can be dramatically reduced if the APE-Gen step is executed on a cluster. For instance, on a machine with 64 threads, with an MHCflurry threshold at 500 nM or 50,000 nM, the same workflow could be executed in 5 hours or 19 hours respectively (without rescoring). Future updates of HLA-Arena should provide additional resources for running workflows in a remote high-performance computing cluster.

HLA-Arena installation

1. If you don't already have it, install Docker.

   Docker for Mac or Windows: https://www.docker.com/products/docker-desktop

   Docker for Linux: https://docs.docker.com/install

2. In a command prompt, pull the HLA-Arena image from Docker Hub by typing:

   docker pull kavrakilab/hla-arena

3. Create a folder in which you want to run the workflows (optional):

   mkdir workflows; cd workflows

4. Copy HLA-Arena notebooks and associated data to your local machine by typing:

   docker run --rm -v $(pwd):/temp --entrypoint cp kavrakilab/hla-arena /hla_arena_data/data.tar.gz \
   /temp/; tar -xzvf data.tar.gz

5. Run HLA-Arena in this folder by typing:

   docker run --rm -v $(pwd):/data -p 8888:8888 --entrypoint="" kavrakilab/hla-arena jupyter \
   notebook --port=8888 --no-browser --ip=0.0.0.0 --allow-root

6. This should generate a URL with the following format:

   http://127.0.0.1:8888/?token=<token_value>

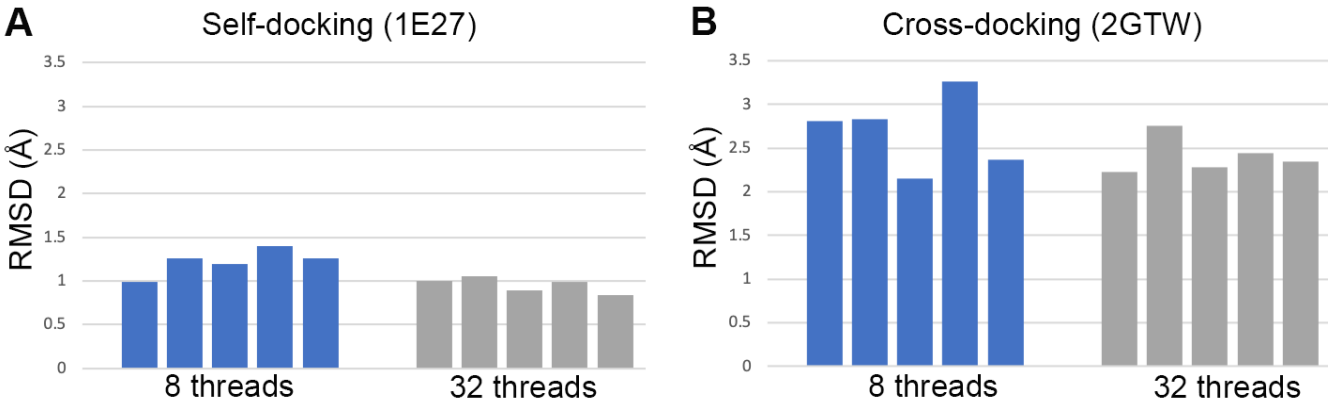
7. Copy and paste this URL into a browser, and open any available Jupyter notebook (i.e., one of the files with extension .ipynb). Note that all the data created in the container will be saved inside sub-directories of the current folder.

8. Check out the file "DOCUMENTATION.html," provided alongside the Jupyter notebooks, for additional information on the workflows and available functions. Enjoy HLA-Arena!

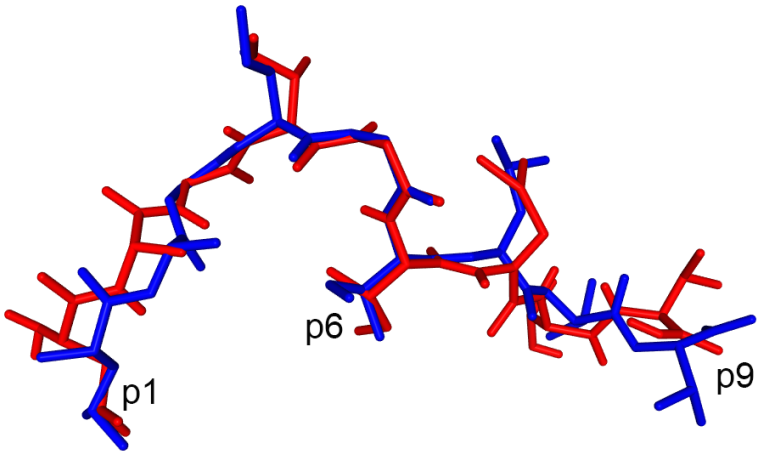
Supplementary Table S 1. Curated dataset of experimentally-determined peptide binders restricted to HLA-A*02:01.

Peptide	Method	Assay	Affinity (nM)	IEDB Label	IEDB ID	PDB ID
FLPSDFFPSV	cellular/radioactivity	IC50	0.57	Positive-High	201178	3OXR
ALWGFFPVL	purified/radioactivity	IC50	2.7	Positive	1775814	1LP9
LLFGYPVYV	purified/radioactivity	KD (≈IC50)	3.8	Positive-High	201486	1DUZ
CINGVCWTV	purified/radioactivity	KD (≈IC50)	55	Positive-Med	1955167	3MRG
VLRDDLLEA	purified/fluorescence	IC50	365	Positive-Low	1809531	3FT4
AAGIGILTV	purified/radioactivity	IC50	395	Positive-Med	201470	3QFD
RQISQDVKL	purified/radioactivity	KD (≈IC50)	1,925	Positive	3243420	4NO5
RGPGRAFVTI	purified/radioactivity	KD (≈IC50)	4,600	Positive-Low	1022278	3ECB
ILKEPVHGV	purified/fluorescence	IC50	7,082	Positive-Med	1783069	2X4U
EAAGIGILTV	purified/fluorescence	IC50	14,560	Positive-Low	2369616	2GT9
SLLMWITQC	purified/radioactivity	KD (≈IC50)	21,070	Positive-Low	208218	2P5E
AAEQRRSTI	cellular/fluorescence	IC50	>70,000	Negative	1873146	–
DAKRNSKSL	cellular/fluorescence	IC50	>70,000	Negative	1872692	–
EIDVSEVKT	cellular/fluorescence	IC50	>70,000	Negative	1874620	–
ATKRYPGVM	cellular/fluorescence	IC50	>70,000	Negative	1875085	–
ETLNEYKQL	cellular/fluorescence	IC50	>70,000	Negative	1873828	–

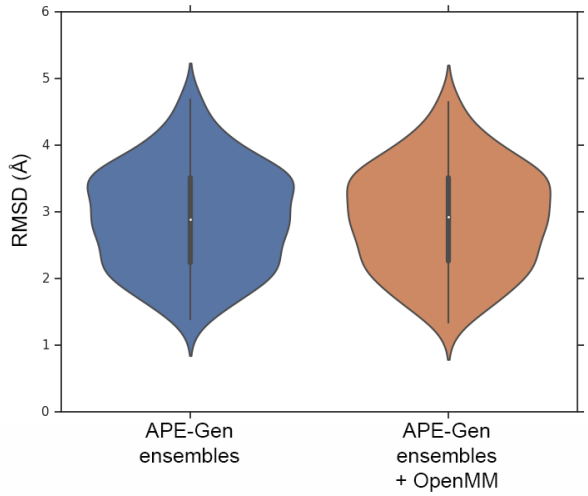
Selected methods include either cellular HLA or purified HLA, used for competitive radioactive or competitive fluorescence measurements. Assays measured either the half maximal inhibitory concentration (IC50) or the dissociation constant (KD). Non-binders are characterized by the lack of a precise measurement in the IEDB (e.g., affinity > 70,000 nM) and of a crystal structure in the PDB.



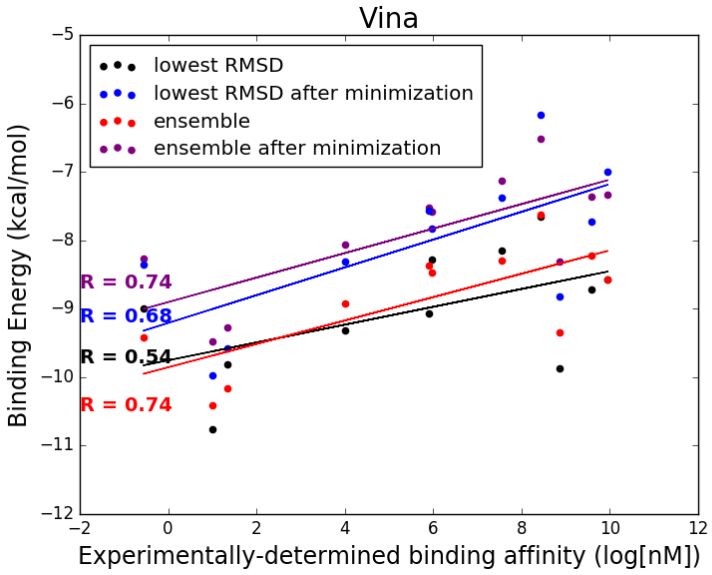
Supplementary Figure S 1. Lowest-RMSD binding modes sampled by DINC in the geometry prediction workflow. **A** Results of a self-docking experiment aimed at reproducing a crystal structure (with PDB code 1E27) involving a 9-mer peptide derived from HIV-1 and the HLA-B*51:01 receptor. This experiment was carried out with either 8 or 32 threads. Each bar corresponds to the so-called *top-RMSD* conformation (i.e., the conformation with the lowest RMSD to the target crystal structure) sampled in each of five replicated runs. Near-native peptide conformations (i.e., conformations with an all-heavy-atom RMSD to the crystal structure less than 2.5 Å) were sampled in all runs. The best conformation across all runs had an all-heavy-atom RMSD of 0.84 Å. **B** Results of a cross-docking experiment aimed at reproducing a crystal structure (with PDB code 2GTW) involving HLA-A*02:01 and a 9-mer peptide derived from the MART-1/Melan-A protein. Near-native peptide conformations were sampled in two out of five runs when using 8 threads, and in four out of five runs when using 32 threads. The best conformation sampled across all runs had an all-heavy-atom RMSD of 2.15 Å.



Supplementary Figure S 2. Lowest-RMSD binding mode sampled in a cross-docking experiment. Depicted in red, is the lowest-RMSD conformation sampled by DINC in the cross-docking experiment aimed at reproducing a crystal structure (with PDB code 2GTW) involving HLA-A*02:01 and a 9-mer peptide derived from the MART-1/Melan-A protein. The all-heavy-atom RMSD of this conformation to the crystal structure (depicted in blue) is of only 2.15 Å. This conformation accurately reproduces the geometry of the first residue (p1), which has an unusual arrangement (i.e., anchored in pocket B of the binding cleft).



Supplementary Figure S 3. Binding mode ensembles generated by APE-Gen include near-native peptide conformations. This plot aggregates the all-heavy-atom RMSD (in Å) between each conformation produced by APE-Gen for each peptide-binder in our dataset (*see* Table S1) and its reference crystal structure. Results for conformations having undergone energy minimization with OpenMM[24] are also reported, although differences are very subtle. These conformations were produced by a single round of sampling with APE-Gen.



Supplementary Figure S 4. Binding energy rankings associated with ensembles or single conformations. This plot reports correlations (assessed as Pearson's R) between experimentally-determined binding affinities and structure-based binding energies predicted by Vina's scoring function using different procedures. More specifically, the binding energy of a given peptide can be defined as: (i) the score of the conformation with the lowest RMSD to the crystal structure in the ensemble produced by APE-Gen (R = 0.54), (ii) the score of that same conformation minimized with OpenMM (R = 0.68), (iii) the median score within the ensemble of conformations produced by APE-Gen (R = 0.74), or (iv) the median score within that same ensemble after minimization with OpenMM (R = 0.74). Each point corresponds to a known peptide-binder to HLA-A*02:01 (*see* Table S1). Note that the non-binders were not included in this analysis.