# DoRIAT:A Bayesian framework for interpreting and annotating docking runs.

**Christos Maniatis** [1] **Zahra Ouaray** [1] **Chengkai Xiao** [1] **Thomas P.E. Dixon** [1] **Charlie Naylor** [1] **James Snowden** [1] **Michelle Teng** [1] **Jacob Hurst** [1]

## Abstract

The advent of sequence-to-structure deep-learning models have transformed protein engineering landscape by providing an accurate and cost effective way to determine crystal structures. Despite their accuracy, deep-learning predictions tend to give limited insights around protein dynamics. To improve conformation exploration we have developed a machine learning pipeline that combines deep-learning predictions with molecular docking. In this report, we propose **Do**cking **R**un **I**ntepretation and **A**nnotation **T**ool (DoRIAT). In contrast to frameworks that score models based on interface interactions, DoRIAT uses a set of parameters that summarize binding conformation. We use DoRIAT to score output from docking runs, identify complexes close to the native structure and create ensembles of models with similar binding conformations. Our results demonstrate that the single structural model DoRIAT selects to be the closest representation of the crystal structure lies within the top 10 of docked models, ranked by Root Mean Squared Distance(RMSD), in around $80\%$ of cases.

## 1. Introduction

T cell-mediated immunity is a crucial aspect of our immune system, playing a vital role in defending against pathogens and cancerous cells. As T cells have the ability to scrutinise the entire proteome within a cell, by interrogating short peptides presented on the surface of cell by human leukocyte antigens (pHLA), they can constitute a very effective means to protect humans from pathogens or other foreign substances (Coles et al., 2020). Characterising the molecular factors governing interactions between T cell receptor (TCR) and pHLA could shed light on T cells' ability to discriminate against antigens (Coles et al., 2020) and pave the way for developing novel T-cell therapies targeting cancer and autoimmune diseases. An example of a complete TCR-pHLA complex from the public domain, determined by X-ray crystallography (Maveyraud & Mourey, 2020), can be found in Fig. 7.

Resolving protein crystals has a long and interesting history dating back to the early 1930's (Bernal & Crowfoot, 1934; Clark & Corrigan, 1932; Kendrew et al., 1960; HODGKIN, 1949). However, the idea that the knowledge of a protein structure could aid in the design of specific ligands, appeared a few years after the launch of the Protein Data Bank, in 1971 (Beddel et al., 1976; Bernstein et al., 1978; Maveyraud & Mourey, 2020). Technological improvements like crystallisation automates, brighter synchrotron X-ray sources, faster detectors, automated structure solution and refinement pipelines have significantly improved time (Dauter & Wlodawer, 2016; Maveyraud & Mourey, 2020), making structural resolution an integral part of pre-clinical target-based drug design. Structure-based design has now delivered drugs for a number of important diseases, including cancer, HIV, glaucoma and hypertension (Souers et al., 2013; Davis et al., 2008). However, end-to-end protein crystal structure determination remains an expensive and laborious process.

The advent of deep-learning based sequence-to-structure models like AlphaFold (Jumper et al., 2021) have transformed the landscape for *insilico* protein engineering. The initial versions of AlphaFold did not accurately model multi-chain complexes, however this has been addressed by a later release (Josh et al., 2024). Although modern sequence-to-structure deep-learning models can generate highly accurate structural predictions, they frequently display a significant degree of structural homogeneity. Therefore, a crucial knowledge gap persists with regards to the incorporation of molecular dynamics and flexibility, as highlighted in (Josh et al., 2024). A single structural model gives a static picture of a system which in reality is in a constant state of flux. An ensemble-based approach can overcome this problem and more comprehensively describe the preserved and transient interactions. Our approach has been to model the various TCR-pHLA complex components with sequence-to-structure models and bring them together with traditional docking algorithms, similar to (Giulini et al., 2023), in a pipeline named MLDock.

[1]Etcembly ltd., Harwell Campus, Oxfordshire, UK. Correspondence to: Jacob Hurst <jacobhurst@etcembly.io>.

More precisely, we model alpha and beta chains of the TCR and the HLA using TCRModel2 (Yin et al., 2023), an AlphaFold derivative. Then HADDOCK3.0 (Dominguez et al., 2003), a physics based docking platform, brings together the TCR, peptide and HLA like similar methods found in literature (Cross et al., 2009; Halperin et al., 2002; Camacho & Vajda, 2002; Smith & Sternberg, 2002). The docking process involves two core steps. During the first step different ligand conformations in the active site of protein are sampled (Halperin et al., 2002). To reduce computational burden, one protein is fixed in space and the second is rotated and translated around the first one (Dominguez et al., 2003). Subsequently, correct docking poses are delineated from incorrect ones with a force-field based scoring function. A summary of the described process is shown in Fig. 6. Each docking run generates hundreds of possible protein conformations which are analyzed to suggest mutations for discovery and engineering of TCRs to create cancer therapeutics (Shafer et al., 2022).

In this report, we present **Do**cking **R**un **I**nterpretation and **A**nnotation **T**ool (DoRIAT) a method for interpreting and annotating the output of physics based docking programs like (Chen & Weng, 2002; Trott & Olson, 2010; Dominguez et al., 2003). DoRIAT is a Gaussian Process (GP) based regression model which associates root mean squared distance (RMSD) between TCR-pHLA docked model and the crystal structure (xtal) using geometric parameters derived from the TCR-pHLA complex. We use DoRIAT to give a favourable scores to conformations which are more likely to initiate immune response (i.e. canonical binders), identify complexes close to the resolved crystal structure and identify similar docking poses to create an ensemble of models with similar binding conformations. This enables a number of downstream use cases that are of deep interest for *insilico* guided protein engineering.

## 2. Methods

DoRIAT[1] implements a Gaussian Process regressor to score MLDock models from six binding mode parameters, three angles and three displacements.

Consider $U$ TCRs for which MLDock can give us $M$ models. For each TCR $u \in (1, \ldots, U)$ and model $m \in (1, \ldots, M)$, $\mathbf{x}_{um}$ describing the binding mode parameters and is given by:

$$\mathbf{x}_{um} = \begin{pmatrix} \theta^c_{um} \\ \theta^t_{um} \\ \theta^r_{um} \\ s^x_{um} \\ s^y_{um} \\ s^z_{um} \end{pmatrix} \tag{1}$$

A detailed description as to how these geometric parameters are obtained can be found in Sec. A.4. The associated response $\mathbf{y}_{um}$ describing the docked model's distance from native structure is modelled as a noisy observation of the function evaluated at $\mathbf{x}_{um}$, $f_{um} = f(\mathbf{x}_{um})$ through likelihood:

$$\mathbf{y}_{um} | f_{um} \sim \mathbf{N}\left(f_{um}, \sigma^2 \mathbf{I}\right) \tag{2}$$

$\mathbb{P}(\mathbf{y}_{um} | f_{um})$ models the data distribution at $\mathbf{x}_{um}$ and the function $f$

$$f \sim \mathcal{GP}(0, \mathscr{K}) \tag{3}$$

where $\mathscr{K}$ is a positive semi-definite function that determines the covariance of $f$ at locations $x_1$ and $x_2$ and reflects our beliefs regarding the behaviour of $f$. In this work we chose the Matérn kernel described by the following formula.

$$\mathscr{K}_{l,\rho}(x_1, x_2) = l^2 \left(1 + \frac{\sqrt{5}\|x_1 - x_2\|_2}{\rho} + \frac{5\|x_1 - x_2\|_2^2}{3\rho^2}\right) \exp\left(-\frac{\sqrt{5}\|x_1 - x_2\|_2}{\rho}\right) \tag{4}$$

where $l$ is the amplitude controlling the marginal variance of the function and $\rho$ is the length-scale controlling spatial variability.

### 2.1. Objective function for selecting models.

Our model selection strategy is a variation of $\mu_{RMSD}$ minimisation which at the same time drastically reduces search space. More precisely, we select the best model is by:

$$\begin{aligned} &\min \mu_{RMSD} \\ &\text{sub. to } \mathrm{CV}_{RMSD} \geq \alpha_{0.95}\ , \end{aligned} \tag{5}$$

where $\alpha_{0.95}$ is the 95$^{\text{th}}$ percentage quantile of $\mathrm{CV}_{RMSD}$ in the docking run. Minimizing the $\mu_{RMSD}$ is there to give the best model and $\mathrm{CV}_{RMSD}$ is a constrain used to reduce the search space while resolving situations where multiple models present very similar $\mu_{RMSD}$. More information can be found in Sec. A.8.

This objective can be problematic when multiple models with identical $\mu_{RMSD}$ present a coefficient of variation above threshold. In practice, we have never encountered an example like that.

### 2.2. Identifying models with similar docking poses using GP covariance.

The covariance function of Eq. 3 not only encodes assumptions about the function $f$, but also dictates the similarity (Rasmussen & Williams, 2018) between docking poses.

[1] A Python implementation of DoRIAT along with data preprocessing and docking data are available at https://zenodo.org/records/14763708

We use that property to build a graph $\mathscr{G}(V, E)$ of vertices $(V)$ and edges $(E)$ representing binding mode parameters and connected poses respectively. The connection between poses is dictated by partial correlation, which measures dependence while accounting for effects of confounding conformations.

Let $f(\mathbf{X}) = \{f_{um} \forall u \in (1, \ldots, U)$ and $m \in (1, \ldots, M)\}$ be the set of RMSD predicted for $U$ TCRs and $M$ models, where $f$ described in Eq. 3, $\mathbf{X} = \{x_{um} \forall u \in (1, \ldots, U)$ and $m \in (1, \ldots, M)\}$. Without loosing generality we chose TCR with index 1 and two conformations $m_1$ and $m_3$. Their predicted RMSDs are given by $f_{1,m_1}$ and $f_{1,m_3}$ respectively. The partial correlation between variables $f_{1m_1}$ and $f_{1m_3}$ is a measure of their conditional association, given the remaining elements $f(\mathbf{X})_{1,-m_1,-m_3}$ of $f(\mathbf{X})$ and it is defined as:

$$\rho_{f_{1m_1}, f_{1m_3} | f(\mathbf{X})_{1,-m_1,-m_3}} = \frac{\mathbf{Cov}\left(f_{1m_1}, f_{1m_3} | f(\mathbf{X})_{1,-m_1,-m_3}\right)}{\sqrt{\mathbf{Var}\left(f_{1m_1} | f(\mathbf{X})_{1,-m_1,-m_3}\right)}\sqrt{\mathbf{Var}\left(f_{1m_3} | f(\mathbf{X})_{1,-m_1,-m_3}\right)}} \quad (6)$$

For the purpose of this project, we assume that any two models with binding mode parameters whose partial correlation is higher than $0.05$ are related. This lays the foundation for choosing similar models and building ensembles around conformations of interest.

### 2.3. Contact Maps

A contact map is a two-dimensional representation used to visualize amino acid interactions between proteins of interest. Each residue in the protein complex is assigned a position on the map and colours are used to indicate interactions. We consider two residues to interact if their $C_A$ atoms are at most $8$ Å apart. We use a binary representation to indicate amino acid interactions, with residues meeting our definition being labelled with $1$ and $0$ otherwise.

In Sec. 3 contact maps are used to asses different ensembles. We separately consider the TCR-HLA and TCR-Peptide contacts as they can give us some idea about interactions that drive affinity and specificity respectively. The contact maps of Fig. 4, 5 are created by averaging across individual contact maps of models in each ensemble.

## 3. Results

DoRIAT is a tool for the exploratory analysis of TCR-pHLA structures produced by docking simulations. It aims to establish a robust docking model scoring system from six binding mode parameters characterising the relative position of the TCR to the pHLA, briefly described in Tab. 1 and more detailed in Sec. A.4. In turn, this is used for a set of downstream tasks such as identify and rank near-native conformations or detect similar conformations. In this paper we will focus on docking outcomes from MLDock, our in-house built docking pipeline, however similar approach could be used to understand simulated conformation from any docking engine.

*Table 1.* **Nomenclature.** Table briefly describing parameters used to summarize TCR-pHLA complexes. Binding mode parameters are bolded. A detailed description of how these parameters are computed can be found in Sec. A.4

| | |
|---|---|
| $v_{rot}$ | Alpha and beta chain rotation symmetry vector pointing towards the HLA. |
| $v_{\text{SB}}$ | Vector between the disulphide bonds' centroids in alpha and beta chain pointing towards beta chain. |
| $p_{\text{TCR}}$ | TCR's centre of mass. |
| $p_{\text{MHC}}$ | MHC's centre of mass. |
| $v_{\text{MHC}_1}$, $v_{\text{MHC}_2}$ | Vectors defining the TCR-pHLA interface plane. |
| $n_{\text{MHC}}$ | Vector perpendicular to $v_{\text{MHC}_1}$ and $v_{\text{MHC}_2}$ pointing towards TCR. |
| $\boldsymbol{\theta^c}$ | Angle between $v_{\text{MHC}_1}$ and $v_{\text{SB}}$. |
| $\boldsymbol{\theta^t}$ | Angle between $v_{\text{MHC}_1}$ and $v_{rot}$. |
| $\boldsymbol{\theta^r}$ | Angle between $v_{\text{MHC}_2}$ and $v_{rot}$. |
| $\boldsymbol{s^x}$ | $v_{\text{MHC}_1}$ direction displacement of $p_{\text{TCR}}$ from $p_{\text{MHC}}$ in the interface plane. |
| $\boldsymbol{s^y}$ | $v_{\text{MHC}_1}$ direction displacement of $p_{\text{TCR}}$ from $p_{\text{MHC}}$ in the interface plane. |
| $\boldsymbol{s^z}$ | Distance of $p_{\text{TCR}}$ projected TCR-pHLA interface plane. |

TCRs can bind in multiple conformations to pHLA. This is also highlighted by the hundreds of models that rank well according to HADDOCK's scoring function, which serves as a proxy for affinity. When TCR-pHLA binding exceeds geometric tolerances, important cell surface interactions do not take place and signaling is disrupted (Adams et al., 2011). We refer to docking geometries that fail to induce activation as non-canonical and we have actively tried to remove them in post-docking analysis. Our first approach for identifying and eliminating non-canonical poses was to derive thresholds from limited publicly available X-ray crystallography structures which can be found in Tab. 2. As the structures used to estimate these ranges presented significant levels of similarity and were different from the systems we were internally interested, their use created challenges in our engineering effort. To overcome the problem of setting geometric limitations, we redefine the problem as a prediction of distance to the crystal structure, where signaling is more likely. We developed DoRIAT, a GP regressor based scoring function that takes into account combinations of binding mode parameters to identify models close to the native structure, schematically described in Fig 6.

Briefly, for each TCR $u$, DoRIAT associates RMSD distance $\mathbf{y}_{um}$ between $C_A$ atoms of docked model $m$ with its corresponding crystal structure with TCR-pHLA binding mode parameters $\mathbf{x}_{um}$ using a GP $f$. Our kernel of choice for the GP is the Matérn with $\nu = \frac{5}{2}$ with amplitude $l$ and length-scale $\rho$. As the RMSD data are initially scaled between $[0, 1]$ and subsequently inverse-transformed by the cumulative cdf of a standard normal, observation noise model is a normal distribution with mean the output of the GP $f$

*Table 2.* **Table summarizing the cut-off thresholds for binding mode parameters.**

| | $\theta^c$ | $\theta^t$ | $\theta^r$ | $s^x$ | $s^y$ | $s^z$ |
|---|---|---|---|---|---|---|
| min | 19 | -25 | -30 | -17 | -7 | 20 |
| max | 75 | 25 | 25 | 12 | 10 | 30 |

and variance $\sigma^2$. More details can be found in Sec. A.4. DoRIAT then constructs a constrained optimization problem involving the predicted RMSD and its corresponding coefficient of variation (CV) to select a model close to the native structure. DoRIAT's covariance matrix is used to identify similar docking conformations allowing for the creation of ensembles around points of interest.

### 3.1. DoRIAT accurately scores output models of docking run.

DoRIAT is trained on docking output of 43 TCRs and subsequently tested on 15 TCRs sourced from STCRDab (Leem et al., 2018). For each TCR the docking platform generates 600 possible conformations of the TCR-pHLA complex. Both training and testing data comprise of a mixture of complexes of a TCR bound to viral or cancer peptides presented by HLA-A*02. For testing we also include proprietary data.

Fig. 1 summarizes the scoring results for docked conformations of HIV (5NMG), viral (5ISZ) and cancer (2PYE) related systems from the test set. DoRIAT accurately predicts docked models far from the native structure, where binding mode parameters range outside the canonical range and can be identified with a thresholding approach as indicated by the orange and blue color-coding. DoRIAT struggles with models closer to the native structure, as indicated by larger deviations in Fig. 1c and 1b. In these examples, some of the docked models present binding mode parameters that are on the edge of what the threshold-based approach would deem canonical. However, there are several examples like 5NMG (Fig. 1a), where such edge cases are absent and DoRIAT is very accurate.

For TCR-pHLA systems like 2PYE and 5ISZ, where DoRIAT's predictions differ from the ground truth, the native complex parameters are listed in the first two lines of Tab 3. For 2PYE and 5ISZ, models like 2PYE_m1 and 5ISZ_m1 have parameters $\theta^t$ and $\theta^c$ close to thresholds presented in Tab 2 and highlight challenges for threshold-based method. In absence of ground truth the presented models could be considered plausible TCR-pHLA complexes, even though their backbone distance from their corresponding native structures is greater than 11.5 Å.While DoRIAT underestimates the models' distance from native structure, it is still able to suggest that all considered models are more than 10 Å away from crystal structure, unlike the threshold based approach which would incorrectly retain them.

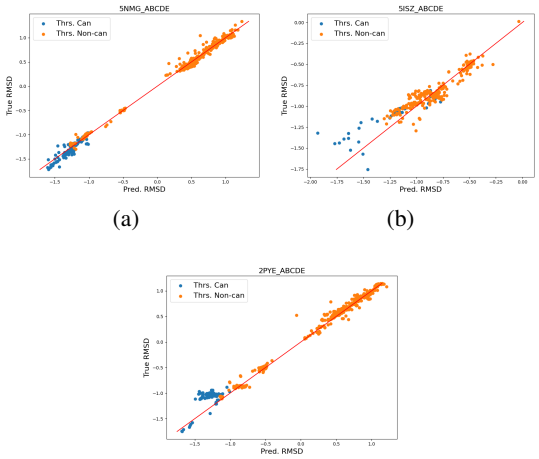



*Figure 1.* **Measured RMSD as a function of predicted RMSD for three TCRs in the test set.** Measured RMSD as a function of predicted RMSD for crystal structures 5NMG (Fig. 1a), 5ISZ (Fig. 1b) and 2PYE (Fig. 1c). Each dot represents a docked model and is color-coded using cut-off thresholds of Tab. 2. The red line corresponds to perfect match between predicted and measured RMSD.

DoRIAT was also applied to docking simulations of unpublished TCR-pHLA complexes, which are more heterogeneous than public data and intended for cancer therapeutics. Their binding mode parameters can be found in rows 3 and 4 of Tab. 3. A direct consequence of the inherent complexity is to have conformations with parameters within range, whose backbone is far from crystal structure causing DoRIAT to struggle with accurate RMSD prediction. TCR A's native structure has $(\theta^c, \theta^r)$ angles of $(40.56°, 10.76°)$, while TCR-pHLA complex A_m1 and TCR-pHLA complex A_m2 bind at $(22.75°, -13.67°)$ and $(70.17°, -16.12°)$. For these examples, the discrepancy between DoRIAT's prediction and actual RMSD would be 5 Å and 3 Å respectively. However, these models would not affect downstream analysis as they would easily be flagged for removal due to their high predicted RMSD (19 Å and 7 Å respectively). TCR B's listed in Tab. 3 (row 4) is one of the most difficult TCRs to make predictions. TCR-pHLB complex B_m1 and TCR-pHLA complex B_m2 are part of a bigger trend presented in Fig. 3b, where DoRIAT underestimates the RMSD from crystal structure as binding mode parameters are far from native structure, but within range.

By jointly evaluating binding mode parameters, DoRIAT is more effective than the threshold-based method, offering higher control over the granularity of the post-docking analysis. However, DoRIAT's training on limited data creates unique challenges in subsequent tasks like finding a model close to the native structure. In following section, we out-

*Table 3.* **Binding mode parameters for section's examined examples.** Summary of the binding poses for native complexes (first four lines) and docking outcomes with canonical parameters that represent structures distant from the ground truth (remaining six lines).

| | $\theta^c$ | $\theta^t$ | $\theta^r$ | $s^x$ | $s^y$ | $s^z$ |
|---|---|---|---|---|---|---|
| 2PYE | 66.73 | 8.21 | -17.03 | 4.39 | -4.96 | 24.74 |
| 5ISZ | 56.33 | 2.66 | -7.34 | 3.34 | 2.71 | 24.53 |
| TCR A | 40.56 | 6.40 | 10.76 | 0.72 | -5.18 | 23.06 |
| TCR B | 65.67 | -3.74 | -8.53 | 2.87 | -0.29 | 25.29 |
| 2PYE_m1 | 41.31 | 23.97 | 8.88 | 4.66 | 1.12 | 22.71 |
| 5ISZ_m1 | 21.75 | 19.46 | 5.81 | 11.48 | -1.04 | 22.14 |
| TCR-pHLA complex A_m1 | 22.75 | 22.49 | -13.67 | 9.75 | -6.61 | 23.08 |
| TCR-pHLA complex A_m2 | 70.17 | -4.87 | -16.12 | 6.02 | -6.93 | 23.59 |
| TCR-pHLA complex B_m1 | 21.55 | 7.06 | 9.32 | 6.69 | -0.53 | 24.08 |
| TCR-pHLA complex B_m2 | 45.56 | -18.38 | -6.85 | -2.42 | -6.46 | 24.85 |

line an approach that leverages the model to address this challenge.

### 3.2. DoRIAT outperforms state of the art in identifying models close to the native structure.

By benchmarking MLDock on public and internal TCR-pHLA complexes, we have identified that the median RMSD between the best simulated pose and the crystal structure is 3.1 Å. This can become particularly useful when structurally assessing candidates from our screening for therapeutics pipeline where access to a crystal structure is not available. Having a method that can suggest a docked models close to the native structure allows to cost-effectively evaluate if a candidate has engineering potential.

The comparison between different model selection approaches is performed using 18 TCR-pHLA complexes, 15 from DoRIAT's test set and 3 internal which include TCR-pHLA complex A and TCR-pHLA complex B. Internal structures are cancer related, while public structures contain a mixture of viral and cancer. All structures are bound to HLA-A$^*$02.

To evaluate the effectiveness of different methods in identifying models close to the native structure, we consider, alongside DoRIAT, MLDock score (HADDOCK score in water refinement), three DeepRank variations and GNN-DOVE. HADDOCK score is a weighted sum of electrostatic, van der Waals, desolvation and distance restraints energy functions to discriminate native looking structures from the rest. Hence, the lower the score the closer to the native structure should be. For DeepRank we train a naive Graph Neural Network (naive-GNN) and DeepRank-GNN (Réau et al., 2023) to predict RMSD using the code provided in DeepRank2. For DeepRank-GNN-esm (Xu & Bonvin, 2024) and GNN-DOVE, we use the pre-trained models provided in DeepRank-GNN-esm and GNN-DOVE to score models.

All models are set to their default hyper-parameters and are trained on the same data split as DoRIAT using Adam optimizer with 5 step patience early stopping policy for a maximum of 20 steps. For the naive GNN and DeepRank-GNN the computationally intensive PSSM step is skipped.

For each of the compared methods, one docked conformation is selected based on each method's scoring approach. For DoRIAT the model is selected such that it solves Eq. 5. Model selection on DeepRank variations naive-GNN and DeepRank-GNN is performed based on predicted RMSD. For DeepRank-GNN-esm and GNN-DOVE selection is based on fraction of native contacts and the probability that the docking decoy has a CAPRI acceptable quality respectively. For MLDock selection is based on lowest HADDOCK score. These predictions are then ranked based on their RMSD from crystal structure. Subsequently, we define the hitrate as a binary variable specific to each method which is 1 for the selected position and 0 otherwise.

Fig. 2 summarizes the cumulative hitrate as a function of ranking. The structural model that DoRIAT selects to be the closest representation of the crystal structure, out-performs compared methods by being within top 10 of docked models, ranked by RMSD, 78% of times (14 out of 18) with 4 times identifying the best model. In comparison, Naive GNN and MLDock identify models within the top 70 in 60% and 40% of cases respectively, while DeepRank-GNN, DeepRank-GNN-esm, and GNN-DOVE achieve only 22%, 17%, and 17%. More results can be found in Tab. 5. Since the docking engine effectively explores the space of conformations and creates plausible interfaces, predicting canonical models from local interactions becomes challenging. Hence, using binding mode parameters gives DoRIAT significant benefit over alternatives. DeepRank's performance further degrades due to the lack of PSSM, which was ignored as it would make unfeasible the assessment of large TCR batchess. The DeepRank-GNN-esm and GNN-DOVE have been pre-trained on antibody-antigen docking results, which differ from our TCR-pHLA systems, therefore performance is not optimal. However, if we were to retrain or fine-tune them on DoRIAT's training data we would expect them to face some of the challenges encountered by the other DeepRank variations. MLDock's scoring function is the third highest performer with the exception of TCR-pHLA complex A where it performs better even compared to DoRIAT.

To further assess the quality of selections, we consider DockQ score (Basu & Wallner, 2016) which incorporates additional measures like $F_{\text{nat}}$, LRMS and iRMS standardized by CAPRI (Lensink & Wodak, 2013). Tab. 4 and Tab. 6 summarize the DockQ scores for model selection in benchmarking across test set and internal TCRs. DoRIAT presents superior performance with all but one selection having at least acceptable quality and an average DockQ score of 0.42. The second best performer is the naive-GNN

*Table 4.* **Assessment of model selections using DockQ score for test set and internal TCRs. (Part1)** Best perfroming selection is marked with bold. Cells are color-coded to indicate whether selection is canonical (green) or not (red) according to the thresholds of Tab. 2. DockQ scores are in the $[0, 1]$ range.

| | DoRIAT | DeepRank (naive-GNN) | DeepRank-GNN | DeepRank-GNN-esm | GNN-DOVE | MLDock |
|---|---|---|---|---|---|---|
| TCR-pHLA complex A | 0.34 | 0.16 | 0.1 | 0.20 | 0.12 | **0.46** |
| TCR-pHLA complex B | **0.24** | 0.05 | 0.06 | 0.04 | 0.07 | 0.11 |
| 2PYE | **0.54** | 0.11 | 0.03 | 0.02 | 0.02 | 0.02 |
| 5ISZ | **0.49** | 0.13 | 0.04 | 0.28 | 0.29 | 0.22 |
| 5NMG | **0.41** | 0.25 | 0.24 | 0.03 | 0.02 | 0.25 |

DeepRank variation with 5 out of 18 selected models having acceptable quality and an average DockQ score of 0.24. The rest DeepRank based models, MLDock and GNN-DOVE have 7 or fewer acceptable models and an average score of less than 0.20. Furthermore, some of the selections made by DeepRank variations, GNN-DOVE and MLDock are not guaranteed to be within canonical range.

In the left-out set, DoRIAT is sometimes outperformed by other methods based on DockQ scores, such as in TCR-pHLA complexes A, TCR-pHLA complex C, 3QFJ and 4MNQ. For the first two , MLDock's selections are slightly better. However, TCR-pHLA complex C has x-axis and y-axis displacements $(s^x, s^y)$ of $(13.1\ \text{Å}, -9.4\ \text{Å})$, which is outside of canonical range. For 3QFJ and 4MNQ naive-GNN method performs best. In all these examples, the main reason other methods get better DockQ scores is attributed to better interface quality. This is expected as DoRIAT focuses on the global geometric properties of the complex. Even in these cases, DoRIAT's selections fall into the same CAPRI quality class with top performers. Overall, DoRIAT exhibits greater consistency in the quality of selections.

In presence of larger training dataset, we would expect the disparity between DoRIAT, naive-GNN and DeepRank-GNN performance on post-docking analysis to close. However, DoRIAT alternatives would struggle as modern docking engines are effective at creating plausible interfaces even for non-canonical conformations. Hence, in this setting, binding mode parameters provide a robust alternative for building machine learning models without the need for massive amounts of data.

### 3.3. DoRIAT covariance can help resolve the post-docking landscape.

Fig. 2 demonstrates that DoRIAT often selects models close to the crystal structure. Beyond that, DoRIAT can also discriminate between poses, giving clear distinction between disconnected conformations and those with similar global geometry, enabling the creation of model ensembles around points of interest. This expands the limited perspective offered by crystal structures or individual models. In this section, we briefly overview how DoRIAT's covariance is used to create ensembles around points of interest using two complexes, TCR-pHLA complex A and TCR-pHLA complex B.

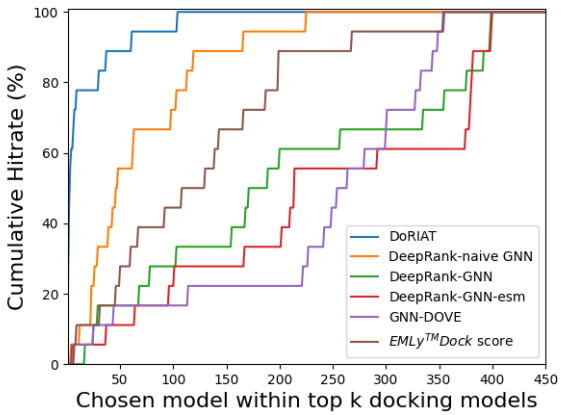



*Figure 2.* **Cumulative hitrate across compared methods.** Graph summarizing cumulative hitrate for DoRIAT (blue), naive GNN (orange), DeepRank-GNN (green), DeepRank-GNN-esm(red) and MLDock (purple) across unseen complexes. Each method chooses one docked model that it believes is closest to the crystal structure. These picks are then ranked using the measured RMSD from crystal structure. We set the lowest ranking to 450 as all methods reach 100% cumulative rate before this position.

Fig. 3c and 3d show PCA of binding mode parameters for complexes A and B, with edges connecting similar conformations. In both examples docking poses cluster around distinct canonical and non-canonical poses. For complex A conformations within canonical range have $\theta^c$ of about $35°$ and $45°$, while for complex B canonical conformation cluster around $\theta^c$ of about $24°$ and $65°$. In absence of a crystal, an ensemble could be created either by considering all canonical models or by picking a model of interest and building the ensemble around it.

To illustrate how covariance information could aid in a more targeted analysis, we create two ensembles of models and compare their average contact maps to those of the crystal structure. The first ensemble includes models which are deemed canonical based on Tab. 2 thresholds. The second uses the optimal model from Eq. 5 along with similar models identified via partial correlation coefficient. Ensemble quality is assessed by comparing TCR-HLA and TCR-peptide contact maps to crystal contact maps using mean structural similarity index (SSIM) (Wang et al., 2004). SSIM is better suited for this task, compared to other distance metrics such as mean squared error, as it accounts for contact position. Proximity to crystal contacts is more tolerable, reflecting protein flexibility, while distant deviations are penalized more heavily as they likely indicate modeling

artifacts.

Fig. 4 and 5 summarize TCR-HLA and TCR-peptide contact maps for the two ensemble methods. In Fig. 4a, both the threshold and optimal model threshold introduce artificial contacts compared to the crystal structure, notably at $(\text{TCR}, \text{HLA}) = (25, 52)$. At the same time, both ensembles identify many correct contacts, resulting in SSIM scores of $0.979$ and $0.975$, with threshold-based ensemble being slightly better. For the TCR-epitope contact map (Fig. 5a), the optimal model ensemble achieves an SSIM $0.815$ against $0.805$ for the threshold-based method, due to fewer contacts in incorrect epitope regions. Fig. 4b suggests that for complex B the ensemble around optimal model matches the crystal contacts, while the threshold-based approach presents erroneous contacts, such as $(\text{TCR}, \text{HLA}) = (26, 55), (204, 150)$. In this case, SSIM for the TCR-HLA contact map of the threshold-based ensemble is $0.970$, $0.015$ lower than that of the optimal model ensemble. The optimal model ensemble, for TCR-peptide contacts, misses contacts at $(\text{TCR}, \text{Peptide}) = (200, 5)$ and adds incorrect ones at $(\text{TCR}, \text{Peptide}) = (30, 3)$ (Fig. 5b). The threshold-based ensemble correctly identifies contacts at $(\text{TCR}, \text{Peptide}) = (200, 5)$, but introduces contacts in the regions $(\text{TCR}, \text{Peptide}) = (30, 3), (48, 3)$ (Fig. 5b). SSIM for the the optimal model ensemble contact map is $0.850$ against $0.843$ for the threshold-based ensemble.

In summary, our results suggest the covariance information can create ensembles with contact maps of equal quality or better quality than those derived from threshold-based methods on binding mode parameters. This enables more accurate interface analysis, improving downstream analysis.

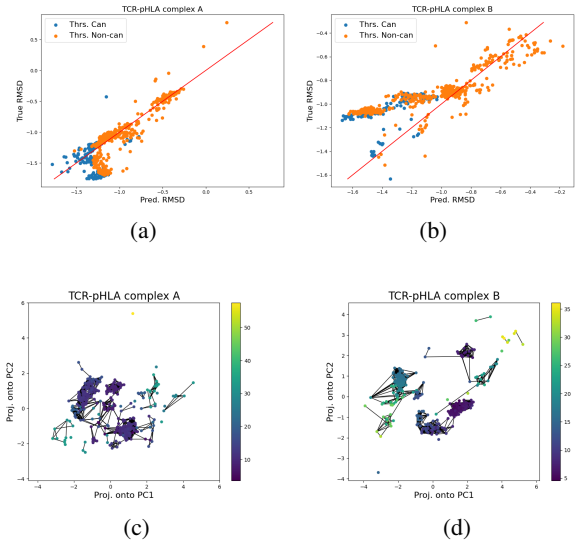


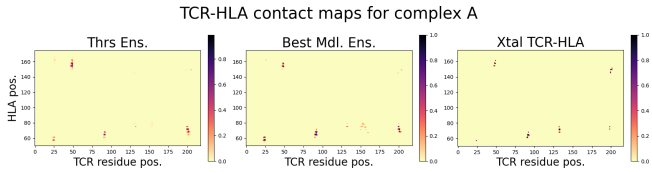

*Figure 3.* **Interpretation of docking runs for internal data.** Figures summarizing post-docking analysis for TCR A and TCR B. (3a,3b) Scatter plots of measured RMSD as a function of predicted RMSD for TCRs that DoRIAT performs well (TCR A) and struggles (TCR B) respectively. Each dot represents a docked model and is color-coded using cut-off thresholds of Tab. 2. The red line represents a perfect match between predicted and measured RMSD. (3c,3d) PCA embedding of binding mode parameters for TCR A and TCR B, color-coded by predicted RMSD. Each dot represents a docked model. Each dot corresponds to a docked model, with edges connecting conformations that share similar binding mode parameters, as determined by partial correlation from GP covariance.

## 4. Conclusion

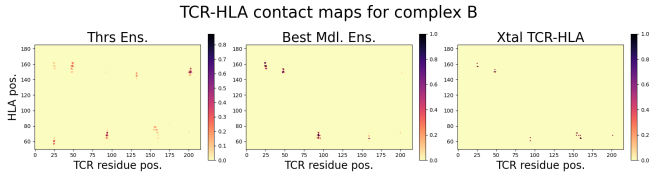
The advent of sequence-to-structure models and their combination with more traditional physics based docking platforms has opened up new possibilities for *insilico* protein engineering. However, traditional docking engines introduce, among native conformations, a series of docking poses which are either impossible to activate the T cell or are far from the native structure observed in X-ray crystallography. We introduce DoRIAT, a Gaussian process regressor which uses binding mode parameters to interpret docking output. By considering combinations of binding mode parameters instead of a threshold-based approach, DoRIAT can reliably generalize on TCR-pHLA complexes of interest.

DoRIAT has been benchmarked in a set of tasks vital for *insilico* protein engineering. One of DoRIAT's standout features is its ability to identify models close to the crystal structure. In contrast to compared methods which heavily rely on interface contacts, DoRIAT's strategy of assessing docked conformations using binding mode parameters allows for accurate and scalable identification of models near to the crystal structure. Additionally, by leveraging DoRIAT's covariance, conformations that exhibit similar binding geometry can be selected creating an ensemble of models around points of interest. These ensembles could be used for a more thorough interface analysis of TCR-pHLA systems of interest. Overall, the described capabilities offer a cost effective solution for evaluating engineering candidates.

DoRIAT has been developed as tool for interpreting ML-Dock's output. Over time it has offered important insights in the effort to engineer picomolar affinity TCRs. We believe that insights provided in this report will offer an alternative way of analyzing docking results which extends beyond TCRs. A similar method to DoRIAT could be developed to analyze antibody-antigen complexes. This is going to come with its own challenges as antibodies lack co-evolutionary constraints, but their large numbers in public repositories could help the algorithm to attain reasonable predictive power.

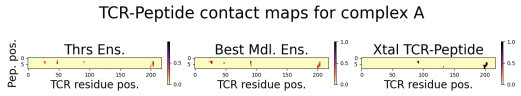



(a)

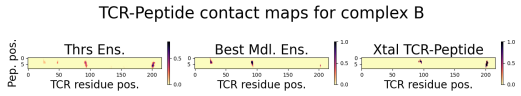



(b)

*Figure 4.* **Contact maps between TCR-HLA** $(8\ \text{Å})$ **for different ensembles for complexes A and B.** Summary of contact maps between the TCR-HLA for complexes A (Fig. 4a) and B (Fig. 4b) for ensembles created using canonical models based on thresholds and similar models around prediction made with Eq. 5. Here a contact is defined if two amino acids have a distance of less than 8 Å. Each contact map is created by averaging contact maps of models in each ensemble. Positions on the contact map are colored based on the average contacts between TCR and HLA. For comparison we include the contact maps observed in the crystal structure.




(a)




(b)

*Figure 5.* **Contact maps between TCR-peptide** $(8\ \text{Å})$ **for different ensembles for complexes A and B.** Summary of contact maps between the TCR-peptide for complexes A (Fig. 5a) and B (Fig. 5b) for ensembles created using canonical models based on thresholds and similar models around prediction made with Eq. 5. Here a contact is defined if two amino acids have a distance of less than 8 Å. Each contact map is created by averaging contact maps of models in each ensemble. Positions on the contact map are colored based on the average contacts between TCR and peptide. For comparison we include the contact maps observed in the crystal structure.

## Impact statement

The development and application of DoRIAT address critical challenges in molecular docking interpretation, enabling a systematic and data-driven approach to analyzing docking results. This tool has the potential to accelerate drug discovery pipelines and improving the identification of high-potential drug candidates. As a result, DoRIAT stands to impact computational chemistry, structural biology and pharmaceutical research communities.

## Lay Summary

T cell-mediated immunity is a cornerstone of the body's defense system, protecting against infections and cancer. Understanding the molecular interactions between T cell receptors (TCRs) and pHLA molecules is key to advancing therapies for cancer and autoimmune diseases. To address this challenge, we have developed a cutting-edge platform that combines deep-learning-based sequence-to-structure models with traditional docking algorithms to analyze TCR-pHLA interactions. At the heart of this effort is DoRIAT, a tool that enables the generation and interpretation of hundreds of potential protein conformations, identifying the most promising candidates for novel T cell therapies.

DoRIAT is a tool for interpreting docking results. It scores and prioritizes conformations most likely to trigger immune responses, identifies models closely resembling experimental crystal structures, and creates ensembles of binding models for more detailed analyses. This unique approach bridges the gap between static protein modeling and the dynamic reality of molecular interactions.

By leveraging this innovative combination of AI and molecular docking, we are positioned to redefine insilico protein engineering. Our solutions offer a scalable, cost-effective way to accelerate the discovery and optimization of T cell therapies, driving us toward a future of transformative treatments for cancer and autoimmune diseases.

## References

Adams, J. J., Narayanan, S., Liu, B., Birnbaum, M. E., Kruse, A. C., Bowerman, N. A., Chen, W., Levin, A. M., Connolly, J. M., Zhu, C., Kranz, D. M., and Garcia, K. C. T cell receptor signaling is limited by docking geometry to peptide-major histocompatibility complex. *Immunity*, 35, 2011. ISSN 10747613. doi: 10.1016/j.immuni.2011.09.013.

Aqvist, J., Luzhkov, V. B., and Brandsdal, B. O. Ligand binding affinities from md simulations. *Accounts of Chemical Research*, 35, 2002. ISSN 00014842. doi: 10.1021/ar010014p.

Basu, S. and Wallner, B. Dockq: A quality measure for protein-protein docking models. *PLoS ONE*, 11, 2016. ISSN 19326203. doi: 10.1371/journal.pone.0161879.

Beddel, C. R., Goodford, P. J., Norrington, F. E., Wilkinson, S., and Wootton, R. Compounds designed to fit a site of known structure in human haemoglobin. *British Journal of Pharmacology*, 57, 1976. ISSN 14765381. doi: 10.1111/j.1476-5381.1976.tb07468.x.

Bernal, J. D. and Crowfoot, D. X-ray photographs of crystalline pepsin [3]. *Nature*, 133, 1934. ISSN 00280836. doi: 10.1038/133794b0.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. *Archives of Biochemistry and Biophysics*, 185, 1978. ISSN 10960384. doi: 10.1016/0003-9861(78)90204-7.

Bissantz, C., Folkers, G., and Rognan, D. Protein-based virtual screening of chemical databases. 1. evaluation of different docking/scoring combinations. *Journal of Medicinal Chemistry*, 43, 2000. ISSN 00222623. doi: 10.1021/jm0010441.

Böhm, H. J. Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3d database search programs. *Journal of Computer-Aided Molecular Design*, 12, 1998. ISSN 0920654X. doi: 10.1023/a:1007999920146.

Camacho, C. J. and Vajda, S. Protein-protein association kinetics and protein docking. *Current Opinion in Structural Biology*, 12, 2002. ISSN 0959440X. doi: 10.1016/S0959-440X(02)00286-5.

Carlson, H. A. and Jorgensen, W. L. An extended linear response method for determining free energies of hydration. *Journal of Physical Chemistry*, 99, 1995. ISSN 00223654. doi: 10.1021/j100026a034.

Charifson, P. S., Corkery, J. J., Murcko, M. A., and Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *Journal of Medicinal Chemistry*, 42, 1999. ISSN 00222623. doi: 10.1021/jm990352k.

Chen, R. and Weng, Z. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins: Structure, Function and Genetics*, 47, 2002. ISSN 08873585. doi: 10.1002/prot.10092.

Clark, G. L. and Corrigan, K. E. The crystal structure of insulin [8]. *Physical Review*, 40, 1932. ISSN 0031899X. doi: 10.1103/PhysRev.40.639.

Coles, C. H., McMurran, C., Lloyd, A., Hock, M., Hibbert, L., Raman, M. C., Hayes, C., Lupardus, P., Cole, D. K., and Harper, S. T cell receptor interactions with human leukocyte antigen govern indirect peptide selectivity for the cancer testis antigen mage-a4. *Journal of Biological Chemistry*, 295, 2020. ISSN 1083351X. doi: 10.1074/jbc.RA120.014016.

Cross, J. B., Thompson, D. C., Rai, B. K., Baber, J. C., Fan, K. Y., Hu, Y., and Humblet, C. Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy. *Journal of Chemical Information and Modeling*, 49, 2009. ISSN 1549960X. doi: 10.1021/ci900056c.

Datta, B. N. Numerical linear algebra and applications. *Numerical Linear Algebra and Applications*, 2010. doi: 10.1137/1.9780898717655.

Dauter, Z. and Wlodawer, A. Progress in protein crystallography. *Protein & Peptide Letters*, 23, 2016. ISSN 09298665. doi: 10.2174/0929866523666160106153524.

Davis, A. M., St-Gallay, S. A., and Kleywegt, G. J. Limitations and lessons in the use of x-ray structural information in drug design. *Drug Discovery Today*, 13, 2008. ISSN 13596446. doi: 10.1016/j.drudis.2008.06.006.

Dillon, J. V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M. D., and Saurous, R. A. Tensorflow distributions. *CoRR*, abs/1711.10604, 2017. URL http://arxiv.org/abs/1711.10604.

Dominguez, C., Boelens, R., and Bonvin, A. M. Haddock: A protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, 125, 2003. ISSN 00027863. doi: 10.1021/ja026939x.

Feher, M. Consensus scoring for protein-ligand interactions. *Drug Discovery Today*, 11, 2006. ISSN 13596446. doi: 10.1016/j.drudis.2006.03.009.

Feher, M., Deretey, E., and Roy, S. Bhb: A simple knowledge-based scoring function to improve the efficiency of database screening. *Journal of Chemical Information and Computer Sciences*, 43, 2003. ISSN 00952338. doi: 10.1021/ci030006i.

Fischer, D., Lin, S. L., Wolfson, H. L., and Nussinov, R. A geometry-based suite of moleculardocking processes. *Journal of Molecular Biology*, 248, 1995. ISSN 00222836. doi: 10.1016/S0022-2836(95)80063-8.

Gehlhaar, D. K., Verkhivker, G. M., Rejto, P. A., Sherman, C. J., Fogel, D. R., Fogel, L. J., and Freer, S. T. Molecular recognition of the inhibitor ag-1343 by hiv-1

protease: conformationally flexible docking by evolutionary programming. *Chemistry and Biology*, 2, 1995. ISSN 10745521. doi: 10.1016/1074-5521(95)90050-0.

Giulini, M., Schneider, C., Cutting, D., Desai, N., Deane, C. M., and Bonvin, A. M. Towards the accurate modelling of antibody-antigen complexes from sequence using machine learning and information-driven docking. *bioRxiv*, 2023.

Halperin, I., Ma, B., Wolfson, H., and Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Structure, Function and Genetics*, 47, 2002. ISSN 08873585. doi: 10.1002/prot.10115.

Hartmann, M. Extending owen's integral table and a new multivariate bernoulli distribution. *arXiv preprint arXiv:1704.04736*, 2017.

Head, R. D., Smythe, M. L., Oprea, T. I., Waller, C. L., Green, S. M., and Marshall, G. R. Validate: A new method for the receptor-based prediction of binding affinities of novel ligands. *Journal of the American Chemical Society*, 118, 1996. ISSN 00027863. doi: 10.1021/ja9539002.

HODGKIN, D. C. The x-ray analysis of the structure of penicillin. *Advancement of science*, 6, 1949. ISSN 0001866X.

Honorato, R. V., Koukos, P. I., Jiménez-García, B., Tsaregorodtsev, A., Verlato, M., Giachetti, A., Rosato, A., and Bonvin, A. M. Structural biology in the clouds: The wenmr-eosc ecosystem. *Frontiers in Molecular Biosciences*, 8, 2021. ISSN 2296889X. doi: 10.3389/fmolb.2021.729513.

Ishchenko, A. V. and Shakhnovich, E. I. Small molecule growth 2001 (smog2001): An improved knowledge-based scoring function for protein-ligand interactions. *Journal of Medicinal Chemistry*, 45, 2002. ISSN 00222623. doi: 10.1021/jm0105833.

Jackson, R. M. Q-fit: A probabilistic method for docking molecular fragments by sampling low energy conformational space. *Journal of Computer-Aided Molecular Design*, 16, 2002. ISSN 0920654X. doi: 10.1023/A:1016307520660.

Jain, A. N. Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities. *Journal of Computer-Aided Molecular Design*, 10, 1996. ISSN 0920654X. doi: 10.1007/BF00124474.

Josh, A., Jonas, A., Jack, D., Richard, E., Tim, G., Alexander, P., Olaf, R., Lindsay, W., J, B. A., Joshua, B., W., B. S., Evans, D. A., Chia-Chun, H., Michael, O., David, R., Kathryn, T., Zachary, W., Žemgulytė Akvilė, Eirini, A., Charles, B., Ottavia, B., Alex, B., Alexey, C., Miles, C., I., C.-R. A., Andrew, C., Michael, F., B., F. F., Hannah, G., Rishub, J., A., K. Y., R., L. C. M., Kuba, P., Anna, P., Pascal, S., Sukhdeep, S., Adrian, S., Ashok, T., Catherine, T., Sergei, Y., D., Z. E., Michal, Z., Žídek Augustin, Victor, B., Pushmeet, K., Max, J., Demis, H., and M., J. J. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630, 2024. ISSN 14764687. doi: 10.1038/s41586-024-07487-w.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. Highly accurate protein structure prediction with alphafold. *Nature*, 596, 2021. ISSN 14764687. doi: 10.1038/s41586-021-03819-2.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., and Shore, V. C. Structure of myoglobin: A three-dimensional fourier synthesis at 2 . resolution. *Nature*, 185, 1960. ISSN 00280836. doi: 10.1038/185422a0.

Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews Drug Discovery*, 3, 2004. ISSN 14741776. doi: 10.1038/nrd1549.

Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews*, 93, 1993. ISSN 15206890. doi: 10.1021/cr00023a004.

Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts of Chemical Research*, 33, 2000. ISSN 00014842. doi: 10.1021/ar000033j.

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., and Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, 161, 1982. ISSN 00222836. doi: 10.1016/0022-2836(82)90153-X.

Leem, J., Oliveira, S. H. D., Krawczyk, K., and Deane, C. M. Stcrdab: The structural t-cell receptor database. *Nucleic Acids Research*, 46, 2018. ISSN 13624962. doi: 10.1093/nar/gkx971.

Lefranc, M. P. Unique database numbering system for immunogenetic analysis; current literature. *Immunology Today*, 18, 1997. ISSN 01675699. doi: 10.1016/S0167-5699(97)01163-8.

Lefranc, M. P. Imgt locus on focus. a new section of experimental and clinical immunogenetics. *Experimental and Clinical Immunogenetics*, 15, 1998. ISSN 02549670. doi: 10.1159/000019049.

Lefranc, M. P. The imgt unique numbering for immunoglobulins, t-cell receptors, and ig-like domains. *Immunologist*, 7, 1999. ISSN 11925612.

Lensink, M. F. and Wodak, S. J. Docking, scoring, and affinity prediction in capri. *Proteins: Structure, Function and Bioinformatics*, 81, 2013. ISSN 08873585. doi: 10.1002/prot.24428.

Maveyraud, L. and Mourey, L. Protein x-ray crystallography and drug discovery. *Molecules*, 25, 2020. ISSN 14203049. doi: 10.3390/molecules25051030.

Meng, X.-Y., Zhang, H.-X., Mezei, M., and Cui, M. Molecular docking: A powerful approach for structure-based drug discovery. *Current Computer Aided-Drug Design*, 7, 2012. ISSN 15734099. doi: 10.2174/157340911795677602.

Meng, Y., Dashti, D. S., and Roitberg, A. E. Computing alchemical free energy differences with hamiltonian replica exchange molecular dynamics (h-remd) simulations. *Journal of Chemical Theory and Computation*, 7, 2011. ISSN 15499618. doi: 10.1021/ct200153u.

Miller, B. R., McGee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. Mmpbsa.py: An efficient program for end-state free energy calculations. *Journal of Chemical Theory and Computation*, 8, 2012. ISSN 15499626. doi: 10.1021/ct300418h.

Mitchell, J. B., Laskowski, R. A., Alex, A., and Thornton, J. M. Bleep - potential of mean force describing protein-ligand interactions: I. generating potential. *Journal of Computational Chemistry*, 20, 1999. ISSN 01928651. doi: 10.1002/(SICI)1096-987X(199908)20:11⟨1165::AID-JCC7⟩3.0.CO;2-A.

Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., and Corbeil, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *British Journal of Pharmacology*, 153, 2008. ISSN 00071188. doi: 10.1038/sj.bjp.0707515.

Muegge, I. and Martin, Y. C. A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *Journal of Medicinal Chemistry*, 42, 1999. ISSN 00222623. doi: 10.1021/jm980536j.

Norel, R., Fischer, D., Wolfson, H. J., and Nussinov, R. Molecular surface recognition by a computer vision-based technique. *Protein Engineering, Design and Selection*, 7, 1994a. ISSN 17410126. doi: 10.1093/protein/7.1.39.

Norel, R., Lin, S. L., Wolfson, H. J., and Nussinov, R. Shape complementarity at protein–protein interfaces. *Biopolymers*, 34, 1994b. ISSN 10970282. doi: 10.1002/bip.360340711.

Norel, R., Petrey, D., Wolfson, H. J., and Nussinov, R. Examination of shape complementarity in docking of unbound proteins. *Proteins: Structure, Function and Genetics*, 36, 1999a. ISSN 08873585. doi: 10.1002/(SICI)1097-0134(19990815)36:3⟨307::AID-PROT5⟩3.0.CO;2-R.

Norel, R., Wolfson, H. J., and Nussinov, R. Small molecule recognition: Solid angles surface representation and molecular shape complementarity. *Combinatorial Chemistry & High Throughput Screening*, 2, 1999b. ISSN 13862073. doi: 10.2174/1386207302666220204193837.

Rasmussen, C. E. and Williams, C. K. I. Gaussian processes for machine learning. *Gaussian Processes for Machine Learning*, 2018. doi: 10.7551/mitpress/3206.001.0001.

Renaud, N., Geng, C., Georgievska, S., Ambrosetti, F., Ridder, L., Marzella, D. F., Réau, M. F., Bonvin, A. M., and Xue, L. C. Deeprank: a deep learning framework for data mining 3d protein-protein interfaces. *Nature Communications*, 12, 2021. ISSN 20411723. doi: 10.1038/s41467-021-27396-0.

Rodrigues, J. P., Teixeira, J. M., Trellet, M., and Bonvin, A. M. Pdb-tools: A swiss army knife for molecular structures. *F1000Research*, 7, 2018. ISSN 20461402. doi: 10.12688/f1000research.17456.1.

Rudolph, M. G., Stanfield, R. L., and Wilson, I. A. How tcrs bind mhcs, peptides, and coreceptors. *Annual Review of Immunology*, 24, 2006. ISSN 07320582. doi: 10.1146/annurev.immunol.23.021704.115658.

Réau, M., Renaud, N., Xue, L. C., and Bonvin, A. M. Deeprank-gnn: a graph neural network framework to learn patterns in protein–protein interfaces. *Bioinformatics*, 39, 2023. ISSN 13674811. doi: 10.1093/bioinformatics/btac759.

Shafer, P., Kelly, L. M., and Hoyos, V. Cancer therapy with tcr-engineered t cells: Current strategies, challenges, and prospects. *Frontiers in Immunology*, 13, 2022. ISSN 16643224. doi: 10.3389/fimmu.2022.835762.

Smith, G. R. and Sternberg, M. J. Prediction of protein-protein interactions by docking methods. *Current Opinion in Structural Biology*, 12, 2002. ISSN 0959440X. doi: 10.1016/S0959-440X(02)00285-3.

Souers, A. J., Leverson, J. D., Boghaert, E. R., Ackler, S. L., Catron, N. D., Chen, J., Dayton, B. D., Ding, H., Enschede, S. H., Fairbrother, W. J., Huang, D. C., Hymowitz, S. G., Jin, S., Khaw, S. L., Kovar, P. J., Lam, L. T., Lee, J., Maecker, H. L., Marsh, K. C., Mason, K. D., Mitten, M. J., Nimmer, P. M., Oleksijew, A., Park, C. H., Park, C. M., Phillips, D. C., Roberts, A. W., Sampath, D., Seymour, J. F., Smith, M. L., Sullivan, G. M., Tahir, S. K., Tse, C., Wendt, M. D., Xiao, Y., Xue, J. C., Zhang, H., Humerickhouse, R. A., Rosenberg, S. H., and Elmore, S. W. Abt-199, a potent and selective bcl-2 inhibitor, achieves antitumor activity while sparing platelets. *Nature Medicine*, 19, 2013. ISSN 10788956. doi: 10.1038/nm.3048.

Teng, M. K., Smolyar, A., Tse, A. G., Liu, J. H., Liu, J., Hussey, R. E., Nathenson, S. G., Chang, H. C., Reinherz, E. L., and Wang, J. H. Identification of a common docking topology with substantial variation among different tcr-peptide-mhc complexes. *Current Biology*, 8, 1998. ISSN 09609822. doi: 10.1016/s0960-9822(98)70160-5.

Terp, G. E., Johansen, B. N., Christensen, I. T., and Jørgensen, F. S. A new concept for multidimensional selection of ligand conformations (multiselect) and multidimensional scoring (multiscore) of protein-ligand binding affinities. *Journal of Medicinal Chemistry*, 44, 2001. ISSN 00222623. doi: 10.1021/jm001090l.

Trott, O. and Olson, A. J. Autodock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31, 2010. ISSN 0192-8651. doi: 10.1002/jcc.21334.

Verkhivker, G., Appelt, K., Freer, S. T., and Villafranca, J. E. Empirical free energy calculations of ligand-protein crystallographic complexes. i. knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Engineering, Design and Selection*, 8, 1995. ISSN 17410126. doi: 10.1093/protein/8.7.677.

Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Arthurs, S., Colson, A. B., Freer, S. T., Larson, V., Lutty, B. A., Marrone, T., and Rose, P. W. Deciphering common failures in molecular docking of ligand-protein complexes. *Journal of Computer-Aided Molecular Design*, 14, 2000. ISSN 0920654X. doi: 10.1023/A:1008158231558.

Wang, X., Flannery, S. T., and Kihara, D. Protein docking model evaluation by graph neural networks. *Frontiers in Molecular Biosciences*, 8, 2021. ISSN 2296889X. doi: 10.3389/fmolb.2021.647915.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13, 2004. ISSN 10577149. doi: 10.1109/TIP.2003.819861.

Xu, X. and Bonvin, A. M. Deeprank-gnn-esm: A graph neural network for scoring protein-protein models using protein language model. *Bioinformatics Advances*, 4, 2024. ISSN 26350041. doi: 10.1093/bioadv/vbad191.

Yin, R., Ribeiro-Filho, H. V., Lin, V., Gowthaman, R., Cheung, M., and Pierce, B. G. Tcrmodel2: High-resolution modeling of t cell receptor recognition using deep learning. *Nucleic Acids Research*, 51, 2023. ISSN 13624962. doi: 10.1093/nar/gkad356.

Zielenkiewicz, P. and Rabczenko, A. Protein-protein recognition: Method of finding complementary surfaces of interacting proteins. *Journal of Theoretical Biology*, 111, 1984. ISSN 10958541. doi: 10.1016/S0022-5193(84)80193-9.

Zundert, G. C. V., Rodrigues, J. P., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., Melquiond, A. S., Dijk, M. V., Vries, S. J. D., and Bonvin, A. M. The haddock2.2 web server: User-friendly integrative modeling of biomolecular complexes. *Journal of Molecular Biology*, 428, 2016. ISSN 10898638. doi: 10.1016/j.jmb.2015.09.014.

Zwanzig, R. W. High-temperature equation of state by a perturbation method. i. nonpolar gases. *Journal of Chemical Physics*, 22, 1954. ISSN 10897690. doi: 10.1063/1.1740409.

# A. Appendix

## A.1. Schematic representation of the pipeline

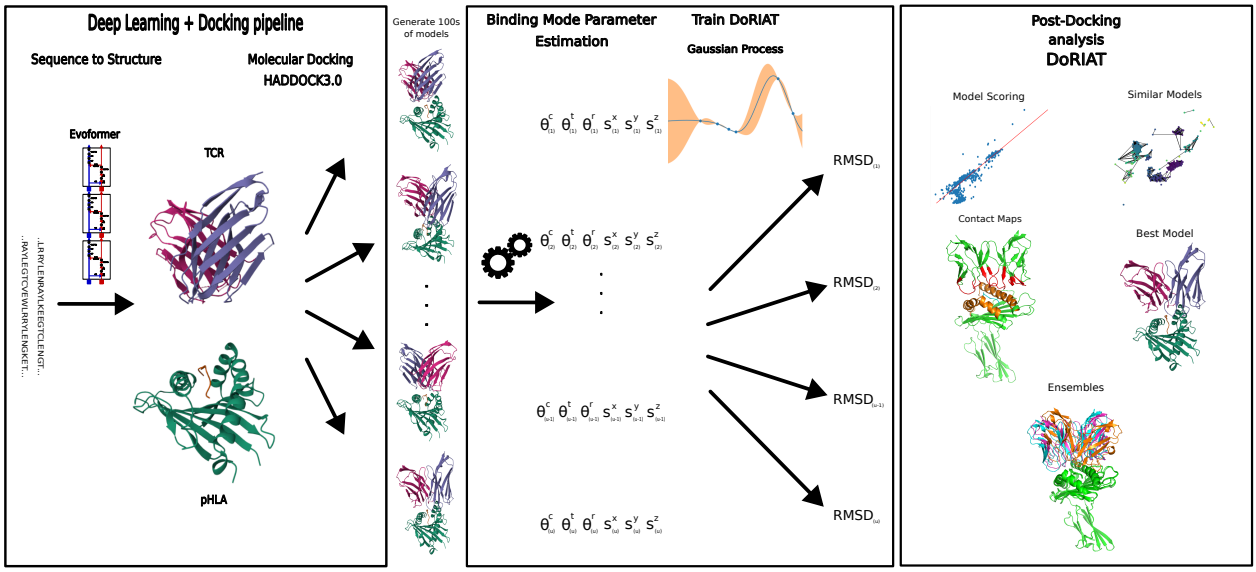



*Figure 6.* **Schematic representation of MLDock and DoRIAT.** In-house developed pipeline that combines deep-learning based sequence-to-structure model to generate alpha and beta chains, followed by *insilico* docking with HADDOCK3.0 (Dominguez et al., 2003) and DoRIAT to interpret docking outcomes and identify the best conformation.

## A.2. Related Work

The binding of proteins and other molecules is fundamental to numerous biological processes. However, predicting binding affinity from structural information alone, within reasonable time, is a challenging task, particularly when experimental data is limited or unavailable. Field practitioners have developed a plethora of approaches which are either used as part of the docking process to score conformations, or post-docking to remove some of the spurious conformations that are less likely to initiate immune response. In this section we review these two approaches explaining some of their underlying assumptions.

Docking scoring functions were developed as a proxy of protein-ligand binding affinity and can be divided into geometric matching, force-field-based, empirical and knowledge-based (Kitchen et al., 2004; Meng et al., 2012).

**Geometric matching** between protein and ligand has repeatedly reaffirmed to have a vital role in determining the geometric complex (Zielenkiewicz & Rabczenko, 1984; Kuntz et al., 1982; Halperin et al., 2002), with early docking algorithms heavily capitalising on it (Fischer et al., 1995; Norel et al., 1994a;b; 1999b;a). However, such efforts were quickly abandoned as they struggle to generalise in absence of ground truth introducing many false positives (Halperin et al., 2002).

**Force-field-based** scoring functions (Carlson & Jorgensen, 1995; Aqvist et al., 2002; Kollman, 1993) assess the binding energy by calculating the sum of electrostatics, calculated by a Coulombic formulation and van der Waals interactions described by the Lennard-Jones potential function (Meng et al., 2012). To address the slow computational speed of force-field-based scoring functions, distant amino acid interactions are ignored. To improve upon standard force-field functions a series of extensions involving hydrogen bonds and solvation energy are included (Jackson, 2002; Moitessier et al., 2008).

In **empirical scoring** functions (Head et al., 1996; Jain, 1996; Verkhivker et al., 2000; Gehlhaar et al., 1995; Böhm, 1998), binding energy decomposes into energy components, such as hydrogen bond, ionic interaction, hydrophobic effect and binding entropy (Meng et al., 2012). The contribution of each component is estimated from a set of ligand-protein complexes with measured affinity. Many modern docking engines such as (Dominguez et al., 2003; Chen & Weng, 2002) are equipped with empirical scoring functions, accounting for a plethora of described phenomena that contribute to affinity. The main limitation of empirical scoring functions is that they might fail to generalize beyond the set of training complexes.

**Knowledge-based** scoring functions (Muegge & Martin, 1999; Mitchell et al., 1999; Ishchenko & Shakhnovich, 2002; Feher

et al., 2003; Verkhivker et al., 1995) are derived by looking into the inter-atomic contact frequencies and protein-ligand distances. Their key assumption is that favorable interactions will be more frequently observed (Meng et al., 2012). Similar to empirical-based scoring approaches, knowledge-based ones often fail to generalize outside of their training data.

**Consensus scoring** (Charifson et al., 1999) combines several different scores to assess the docking conformation. It has been reported (Feher, 2006; Charifson et al., 1999; Bissantz et al., 2000; Terp et al., 2001) that consensus scoring improves prediction of bound conformations and poses (Meng et al., 2012). However, binding energies might still be inaccurate and its predictive power depends on the independence of scoring functions it is made of (Kitchen et al., 2004; Feher, 2006; Charifson et al., 1999; Bissantz et al., 2000; Terp et al., 2001).

A limitation shared across the scoring functions described above is their limited treatment of solvation energy. More rigorous approaches for estimating binding affinity would include Free Energy Perturbation (Zwanzig, 1954), Replica Exchange Free Energy Perturbation (Meng et al., 2011), Molecular Mechanics Poisson-Boltzmann solvent-accessible surface area (MM-PB/SA) (Kollman et al., 2000). However, they are computational demanding and can easily become prohibitively expensive for larger molecules(Miller et al., 2012). A work-around would be to use them for scoring post-docking.

The introduction of powerful hardware accelerators like GPUs and TPUs along with parallel file system technologies has given rise to machine learning approaches. Machine learning methods are extremely flexible and in numerous cases have proved their ability to extract patterns from millions of data points. (Wang et al., 2021) have introduced Graph Neural Network–based DOcking decoy eValuation scorE (GNN-DOVE), a method that scores docking models by looking at atom chemical properties and inter-atom distances in the TCR-pHLA interface. Similar to GNN-DOVE, DeepRank and its variations (Renaud et al., 2021; Réau et al., 2023; Xu & Bonvin, 2024) use graph neural networks (GNNs) to learn residue level features and score biomolecular complexes. Subsequently they could be used for binding affinity prediction but also to disentangle crystal artefacts from protein interactions of potential biological interest.

### A.3. Information driven docking

Information-driven docking integrates experimental or computational data to improve the prediction of biomolecular interactions, enhancing accuracy and efficiency (Dominguez et al., 2003). HADDOCK uses Ambiguous Interaction Restraints (AIRs) to handle incomplete or ambiguous data, enabling the modeling of flexible interactions (Zundert et al., 2016). Unlike unambiguous restraints, AIRs allow for multiple possible interactions between residue groups, making them ideal for uncertain data. This flexibility enables HADDOCK to explore a broader range of conformations while still being guided by the available data (Honorato et al., 2021).

To systematically explore the interaction space, we devised three distinct docking protocols, each defining active and passive residues differently. Protocol A a more permissive approach, with only a minimal number of active residues, allowing for a less constrained exploration. In contrast, protocols B and C were designed to be more targeted, by employing a larger set of active residues in the binding interface. This tiered approach enables a balanced exploration of the conformation space, combining the flexibility of Protocol A with the precision of Protocols B and C, thereby capturing both broad and specific aspects of the interaction landscape.

To train DoRIAT, we used all three protocols. For best model selection we choose the best model from subset of models docked with Protocols B and C, but the ranking was done on all three data. The ensembles were created from Protocol B and C models.

### A.4. Binding mode parameters

Binding mode parameters are essential in assessing T cell activation and subsequent immune response. DoRIAT depends on three angles and three displacements. To estimate these parameters, 2 vectors and a centre of mass for the TCR and pHLA have to be characterized respectively. Here we describe the procedure to obtain DoRIAT features and summarize the process into a pseudo-code.

Let $v_{rot}$ be the vector parallel to the alpha and beta chain rotation symmetry axis pointing towards the HLA, $p_{\text{TCR}}$ be the TCR's centre of mass, and $v_{\text{SB}}$ the vector in the direction defined by alpha and beta chain disulphide-bridges pointing towards beta chain. Let $\text{a}_{\text{alpha}}$ and $\text{a}_{\text{beta}}$ be the set of amino acids' carbon alpha ($C_A$) atomic coordinates in alpha and beta chains respectively aligned based on the IMGT numbering (Lefranc, 1997; 1998; 1999). Let $J$ be the indices of subsets of $\text{a}_{\text{alpha}}$ and $\text{a}_{\text{beta}}$ with common IMGT numbering and $\text{a}_{\text{alpha},J}$ and $\text{a}_{\text{beta},J}$ the relevant subsets. We consider only points with common IMGT numbering in order not to tip the centre of mass and symmetry axis towards the beta chains which is

commonly longer. Let $v_{\text{TCR}}$ be the set of midpoints between alpha and beta chain atomic $C_A$ coordinates with the same IMGT numbering.

$$v_{\text{TCR},j} = \frac{\text{a}_{\text{alpha},J(j)} + \text{a}_{\text{beta},J(j)}}{2} \text{ for } j \text{ in } 0, \ldots \|J\| - 1 \tag{7}$$

Then $v_{rot}$ is the normalized vector pointing towards the HLA parallel to the best fit line of $v_{\text{TCR}}$ coordinates and $p_{\text{TCR}}$ is the mean of $v_{\text{TCR}}$ coordinates. $v_{\text{SB}}$ is the normalized vector parallel to the line connecting disulphide bonds in alpha and beta chains, with the vector pointing towards beta chain.

Let $v_{\text{H}_1}$ and $v_{\text{H}_2}$ be the normalized vectors parallel to the best fit lines running through $C_A$ atoms' coordinates in HLA helix 1 and 2 respectively. Helices 1 and 2 are defined to be the parts of HLA with IMGT numbering ranges $52 - 88$ and $141 - 176$ respectively. To make results consistent when estimating $v_{\text{H}_1}$ and $v_{\text{H}_2}$ we make sure the vectors point towards the last residue in the amino acid sequence. $v_{\text{MHC}_1}$ is the normalized bisecting vector of the parallelogram spanned by $v_{\text{H}_1}$ and $v_{\text{H}_2}$. Hence,

$$v_{\text{MHC}_1} = \frac{v_{\text{H}_1} + v_{\text{H}_2}}{\|v_{\text{H}_1} + v_{\text{H}_2}\|_2} \tag{8}$$

Let $p_{\text{H1}}$ and $p_{\text{H2}}$ be the center of mass estimated from $C_A$ atom coordinates of helices 1 and 2 respectively and $v_{\text{H12}}$ the vector parallel to the line defined by points $p_{\text{H1}}, p_{\text{H2}}$ pointing towards helix 2. Then

$$v_{\text{MHC}_2} = \frac{(v_{\text{MHC}_1} \times v_{\text{H12}}) \times v_{\text{MHC}_1}}{\|(v_{\text{MHC}_1} \times v_{\text{H12}}) \times v_{\text{MHC}_1}\|} \tag{9}$$

Where $\times$ is the cross product [2]. The centre of HLA mass $p_{\text{MHC}}$ is defined as

$$p_{\text{MHC}} = \frac{p_{\text{H1}} + p_{\text{H2}}}{2} \tag{10}$$

Using vectors $v_{\text{MHC}_1}, v_{\text{MHC}_2}$ and the HLA centre of mass $p_{\text{MHC}}$ we define the HLA plane. The equation of the HLA plane for any given point $k$ on the plane is given by :

$$n_{\text{MHC}} \cdot k = n_{\text{MHC}} \cdot p_{\text{MHC}} \,, \tag{11}$$

where

$$n_{\text{MHC}} = \frac{v_{\text{MHC}_1} \times v_{\text{MHC}_2}}{\|v_{\text{MHC}_1} \times v_{\text{MHC}_2}\|_2} \tag{12}$$

Given $v_{rot}$,$v_{\text{SB}}$, $p_{\text{TCR}}$ for the TCR, along with $v_{\text{MHC}_1}$, $v_{\text{MHC}_2}$, $p_{\text{MHC}}$ for the HLA and the HLA plane described by Eq 11 we can estimate the DoRIAT parameters. Visualization of $v_{rot}$,$v_{\text{SB}}$, $p_{\text{TCR}}$, $v_{\text{MHC}_1}$, $v_{\text{MHC}_2}$ and $p_{\text{MHC}}$ for the case of the 4FTV crystal structure can be found in Fig. 7. Let $\theta^c$, $\theta^t$ and $\theta^r$ be the cross (Rudolph et al., 2006), tilt(Teng et al., 1998) and roll(Teng et al., 1998) angles. Then

$$\tilde{\theta_c} = \arccos \frac{v_{\text{MHC}_1} \cdot v_{\text{SB}}}{\|v_{\text{MHC}_1}\|_2 \|v_{\text{SB}}\|_2}$$

$$\theta^c = \begin{cases} \tilde{\theta_c}, \text{ if } \tilde{\theta_c} \geq 90^\circ \\ 360^\circ - \arccos \frac{v_{\text{MHC}_1} \cdot v_{\text{SB}}}{\|v_{\text{MHC}_1}\|_2 \|v_{\text{SB}}\|_2}, \text{ otherwise} \end{cases} \tag{13}$$

$$\theta^t = \arccos \frac{v_{\text{MHC}_1} \cdot v_{rot}}{\|v_{\text{MHC}_1}\|_2 \|v_{rot}\|_2} - 90^\circ \tag{14}$$

$$\theta^r = \arccos \frac{v_{\text{MHC}_2} \cdot v_{rot}}{\|v_{\text{MHC}_2}\|_2 \|v_{rot}\|_2} - 90^\circ \tag{15}$$

Let $s^z$ be the $z$ axis displacement of $p_{\text{TCR}}$ from the plane 11. Then

$$s^z = \frac{|n_{\text{MHC}} \cdot p_{\text{TCR}} - n_{\text{MHC}} \cdot p_{\text{MHC}}|}{\|n_{\text{MHC}}\|_2} \tag{16}$$

[2] An alternative way of obtaining Eq. 9 is by starting with vectors $v_{\text{MHC}_1}$ and $v_{\text{H12}}$ and creating an orthonormal vector basis with Gram–Schmidt process (Datta, 2010).

(a) (b)

*Figure 7.* **Vectors and points required to obtain binding mode parameters for 4FTV structure.** Visual depiction of $v_{rot}$(vector in light green), $v_{\text{SG}}$(vector in red), $v_{\text{MHC}_1}$(vector in blue), $v_{\text{MHC}_2}$(vector in cyan), $p_{\text{TCR}}$ (top light green sphere), $p_{\text{MHC}}$ (blue sphere) with (Fig. 7a) and without (Fig. 7b) ribbon representation of the complex overlayed. The bottom light green sphere on the HLA plane is the projection of $p_{\text{TCR}}$ used to estimate the displacements $s^x$,$s^y$ and $s^z$.

**Algorithm 1** Pseudocode for estimating binding mode parameters

**Require:** A PDB file with the TCR-pHLA complex atoms and their corresponding 3D coordinates.
1: Align complex amino acids based on IMGT numbering.
2: Obtain $C_A$ atoms coordinates from alpha and beta chain with common IMGT numbering.
3: Estimate $v_{\text{TCR}}$ from Eq. 7 and obtain $v_{rot}$ and $p_{\text{TCR}}$.
4: Obtain $C_A$ atoms in the HLA and define vectors $v_{\text{H}_1}$, $v_{\text{H}_2}$ along with centre of masses $p_{\text{H}_1}$ and $p_{\text{H}_2}$ for helices 1 and 2.
5: Define vector $v_{\text{SB}}$ to be parallel to the line connecting alpha and beta chain disulphide bonds with a direction towards beta chain.
6: Using Eq. 8 estimate $v_{\text{MHC}_1}$.
7: Define vector $v_{\text{H}_{12}}$ as the vector parallel to the line connecting $p_{\text{H}_1}$ and $p_{\text{H}_2}$ with direction towards helix 2.
8: Using Eq. 9 estimate $v_{\text{MHC}_2}$.
9: Using Eq. 10 and 12 estimate the HLA center of mass and the normal to the HLA plane defined by $v_{\text{MHC}_1}$ and $v_{\text{MHC}_2}$.
10: Cross $(\theta^c)$, tilt $(\theta^t)$ and roll $(\theta^r)$ angles can be estimated from Eq. 13, 14 and 15 respectively.
11: x-axis $(s^x)$, y-axis $(s^y)$ and z-axis $(s^z)$ displacements can be estimated from Eq. 17, 18 and 16 respectively.

Let $s^x$ and $s^y$ be the $v_{\text{MHC}_1}$ and $v_{\text{MHC}_2}$ axis displacements of $p_{\text{TCR}}$ from $p_{\text{MHC}}$ on the plane 11. Then

$$s^x = \frac{(p_{\text{TCR}} - p_{\text{MHC}}) \cdot v_{\text{MHC}_1}}{\|v_{\text{MHC}_1}\|_2} \tag{17}$$

and

$$s^y = \frac{(p_{\text{TCR}} - p_{\text{MHC}}) \cdot v_{\text{MHC}_2}}{\|v_{\text{MHC}_2}\|_2} \tag{18}$$

The procedure to estimate the binding mode parameters is summarized in Alg. 1/

### A.5. Data Preprocessing

The development of DoRIAT, involved a dataset of 58 crystal structures of human TCRs obtained from the Structural T-cell Receptor Database (STCRDab)(Leem et al., 2018) [3] filtered to keep only ones that bind to HLA-A*02. Using pdb-tools (Rodrigues et al., 2018) we remove all the non-protein atoms such as water, ions, other ligands for the TCR-pHLA complex. Then alpha and beta chains were extracted into fasta format. Subsequently TCR alpha and beta chains are modeled using sequence-to-structure deep-learning based models.

Using MLDock we obtain 600 possible conformation of how the TCR-pHLA complex looks. From these docked structures, we calculate six parameters summarizing the global mode of TCR binding to the pHLA: cross, tilt, roll, shiftx, shifty, and shiftz. More information about these parameters can be found in Section A.4. To quantify the accuracy of the docked structures, we calculated the backbone distances from crystal using only $C_A$ atoms. Crystal structures with missing residues were ignored. Subsequently, backbone distances were normalized to a range between 0 and 1 by dividing with the largest observed distance plus 5 between the docked models and corresponding crystal structures. The normalized distance ranges

[3] Until November 2021.

were then converted to $(-\infty, \infty)$ range to enable the seamless use of GP regressor from Tensorflow probability (Dillon et al., 2017). For model training and testing, we split the data based on complex id (TCR-pHLA) labels, maintaining a 3 to 1 ratio respectively.

As TCR-pHLA complexes contain four amino acid chains and the deep-learning methods examined are designed to score complexes containing only two chains, pdb files are preprocessed accordingly by merging alpha with beta chains and HLA with peptide in the described order.

### A.6. Implementation

For a training set of independent variables $\mathbf{X} = \{\mathbf{x}_{um} \forall u \in (1, \ldots, U), (1, \ldots, M)\}$ and dependent variables $\mathbf{Y} = \{\mathbf{y}_{um} \forall u \in (1, \ldots, U), (1, \ldots, M)\}$. To tune the GP kernel parameters $l$ and $\rho$ of Eq. 4 over training data we minimise the negative log marginal likelihood

$$\mathrm{NL}(\mathbf{Y}) = -\ln\left[\int \mathbf{N}\left(f, \sigma^2\mathbf{I}\right)\mathbf{N}\left(f|0, \mathscr{K}\left(\mathbf{X}, \mathbf{X}\right)\right) df\right], \tag{19}$$

using Adam with learning rate $0.02$. To minimise resource use, we have an early stopping policy applied if the negative log-likelihood drops by less than $1$ unit in $10$ epochs. To prevent over-fitting we tried cross validation but final parameters did not defer much.

### A.7. Mathematical Derivations

DoRIAT is trained and tested with response variable lying in the $(-\infty, \infty)$ range. For predictions and model delineation dependent values have to be converted back into they original Å scale. In this section, we provide the detailed mathematical calculations required to avoid costly simulation along with the justification for using a non-standard approach to select the best model.

Suppose that DoRIAT for some input variable $x^*$ has predicted $\mu^*$ and $\sigma^*$ as latent mean and standard deviation.

Let $\mu_{RMSD}$ denote the RMSD average in observed space. To estimate $\mu_{RMSD}$ we work as follows:

$$\mu_{RMSD} = \int_{-\infty}^{\infty} \epsilon\Phi(x)\mathbf{N}(x, \mu^*, \sigma^*)dx$$

where $\Phi$ is the commulative cdf of standard normal and $\epsilon$ is the maximum observed RMSD we could observe. Using (Hartmann, 2017) we get the following result.

$$\mu_{RMSD} = \int_{-\infty}^{\infty} \epsilon\Phi(x)\mathbf{N}(x, \mu^*, \sigma^*)dx = \epsilon\Phi\left(\frac{\mu^*}{\sqrt{1+\sigma^{*2}}}\right) \tag{20}$$

Let $\sigma_{RMSD}$ denote the RMSD standard deviation in observed space. To estimate $\sigma_{RMSD}$ we work as follows:

$$\sigma_{RMSD} = \sqrt{\int_{-\infty}^{\infty} \epsilon^2\Phi(x)^2\mathbf{N}(x, \mu^*, \sigma^*)dx - \mu_{RMSD}^2}$$

Again using (Hartmann, 2017) we get:

$$\int_{-\infty}^{\infty} \epsilon^2\Phi(x)^2\mathbf{N}(x, \mu^*, \sigma^*)dx = \epsilon^2\mathbf{F}\left(\vec{m}|\vec{0}, V\right)$$

where $\mathbf{F}$ is the bivariate normal cdf with mean $\vec{0}$ and variance $V$ and

$$V = \begin{bmatrix} \sigma^2+1 & \sigma^2 \\ \sigma^2 & \sigma^2+1 \end{bmatrix}, \vec{m} = \begin{bmatrix} \mu^* \\ \mu^* \end{bmatrix}, \vec{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Hence:

$$\sigma_{RMSD} = \epsilon\sqrt{\mathbf{F}\left(\vec{m}|\vec{0}, V\right) - \Phi\left(\frac{\mu^*}{\sqrt{1+\sigma^{*2}}}\right)^2} \tag{21}$$

So, CV is estimated by:

$$\mathrm{CV}_{RMSD} = 100\frac{\sigma_{RMSD}}{\mu_{RMSD}} \tag{22}$$

Predicted entropy $E_{RMSD}$ is given by the following formula:

$$E_{RMSD} = -\int_{-\infty}^{\infty} \Phi(x)\ln\left[\Phi(x)\right]\mathbf{N}(x,\mu^*,\sigma^*)dx$$

Since there is not closed form solution to the integral we first need to approximate $-\Phi(x)\ln\left[\Phi(x)\right]$. A very good approximation can be obtained by Laplace approximation. By doing so we find that the mean is $\mu_L = \Phi^{-1}\left(\frac{1}{e}\right)$ and standard deviation $\sigma_L = e\mathbf{N}\left(\Phi^{-1}\left(\frac{1}{e}\right),0,1\right)$. Using numerical integration on $-\Phi(x)\ln\left[\Phi(x)\right]$ we see that $\int_{-\infty}^{\infty} -\Phi(x)\ln\left[\Phi(x)\right]dx = 0.90319728556$. Hence $c\mathbf{N}\left(x,\mu_L,\sigma_L\right)$ is a really good approximation of $-\Phi(x)\ln\left[\Phi(x)\right]$. Hence:

$$E_{RMSD} \approx \int_{-\infty}^{\infty} c\mathbf{N}(x,\mu_L,\sigma_L)\mathbf{N}(x,\mu^*,\sigma^*)dx$$

$$E_{RMSD} \approx c\mathbf{N}(\mu^*,\mu_L,\sqrt{\sigma^{*2}+\sigma_L^2}) \tag{23}$$

### A.8. Objective function justification

In this section we attempt to provide some additional insights as to why we chose Eq. 5 as our objective function.

To construct an objective function that selects good models we start by examining $\mu_{RMSD}$. Eq. 20 and Fig.8a indicate that if were to chose models based on $\mu^*$ and $\sigma^*$, we would have a hard time as different combinations will yield the same results. Hence, it is important for our objective function to break such ties.

To resolve ties in model selection, we prioritize models with a high $CV_{RMSD}$. In the face of uncertainty, we are making the optimistic assumption that we are overestimating $\mu_{RMSD}$. In practice we have observed that this assumption works well, as models close to the crystal structure present higher coefficient of variation compared to the remaining ones. By applying a constraint that retains the top $5\%$ of models in terms of $CV_{RMSD}$, we effectively filter out less favourable conformations. Then by selecting the model with the lowest $\mu_{RMSD}$ among models within that subset we increase the chance of getting a desirable results.

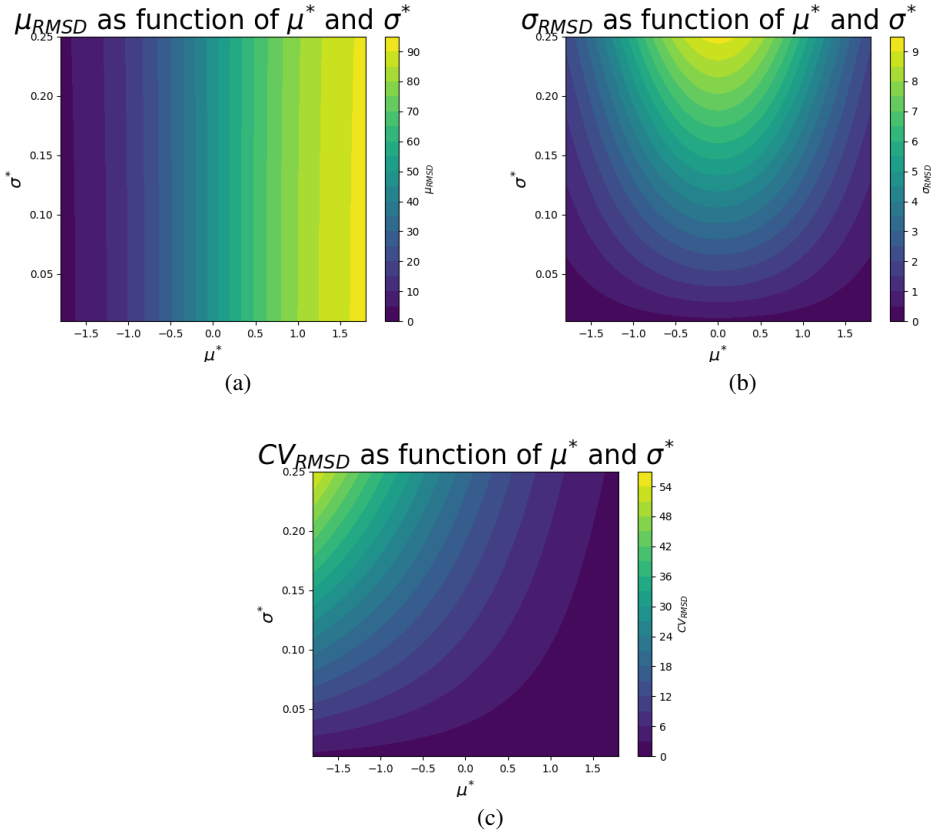



*Figure 8.* **Contour plots of Eq. 20,21 and 22** Contour plots of predicted mean ($\mu_{RMSD}$, Fig 8a), standard-deviation ($\sigma_{RMSD}$, Fig 8b) and coefficient of variation ($\text{CV}_{RMSD}$, Fig 8c) as a function of parameters $\mu^*$ and $\sigma^*$. Thresholds for $\mu^*$ and $\sigma^*$ are estimated from test set observations.

### A.9. Cumulative hitrate across compared methods

In Sec. 3.2 we select a model from the docking output using different scoring approaches. Here we present a table summarizing the cumulative hitrate across compared methods.

*Table 5.* **Table summarizing cumulative hitrate across compared methods.** Table summarizing cumulative hitrate for DoRIAT, naive GNN, DeepRank-GNN, DeepRank-GNN-esm and MLDock across unseen complexes. Each method chooses one docked model that believes it is closest to the crystal structure. These picks are then ranked using the measured RMSD from crystal structure.

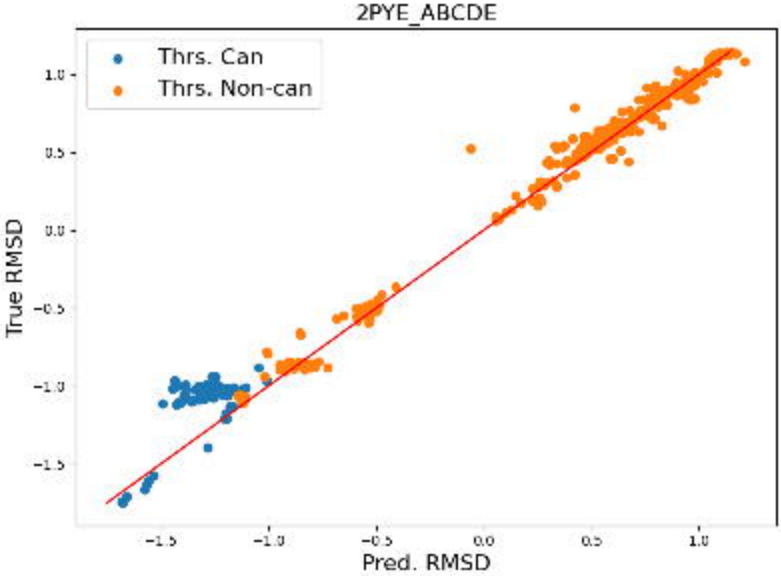
| Hitrate / Top | 1 | 10 | 20 | 40 | 70 | 120 | 170 | 250 |
|---|---|---|---|---|---|---|---|---|
| DoRIAT | 22.22 | 77.78 | 77.78 | 88.89 | 94.44 | 100.00 | 100.00 | 100.00 |
| DeepRank (naive-GNN) | 0.00 | 5.56 | 11.11 | 38.89 | 66.67 | 88.89 | 94.44 | 100.00 |
| DeepRank-GNN | 0.00 | 0.00 | 5.56 | 16.67 | 22.22 | 33.33 | 44.44 | 61.11 |
| DeepRank-GNN-esm | 0.00 | 5.56 | 5.56 | 11.11 | 16.67 | 27.78 | 33.33 | 55.56 |
| GNN-DOVE | 0.00 | 5.56 | 5.56 | 11.11 | 16.67 | 22.22 | 22.22 | 44.44 |
| MLDock | 0.00 | 11.11 | 11.11 | 16.67 | 38.89 | 50.00 | 72.22 | 88.89 |

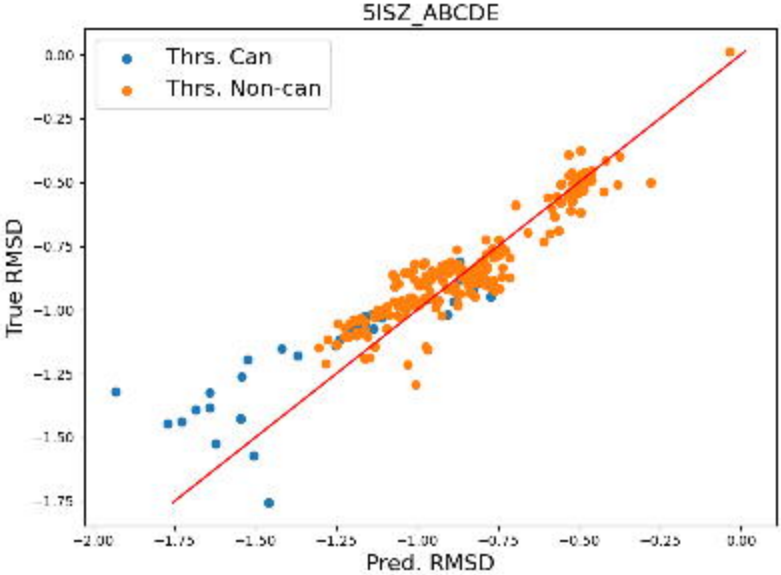
### A.10. Benchmarking model selections with DockQ

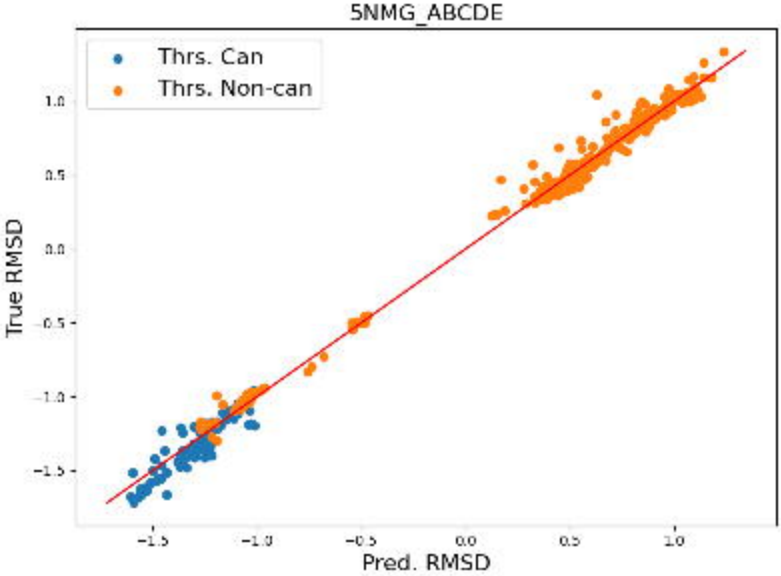
In Sec. 3.2 we select a model from the docking output using different scoring approaches. Apart from ranking predictions based on RMSD, we also compare them using DockQ score. Since DockQ takes into account interface properties apart from ligand RMSD, it provides a more holistic assessment of the docking pose. Here we provide DockQ scores from the remaining test set and internal TCRs.

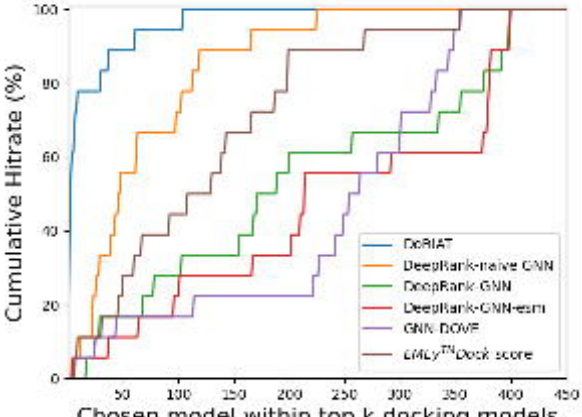
Table 6: **Assessment of model selections using DockQ score for test set and internal TCRs. (Part2)** Best perfroming selection is marked with bold. Cells are color-coded to indicate whether selection is canonical (green) or not (red) according to the thresholds of Tab. 2. DockQ scores are in the $[0, 1]$ range.

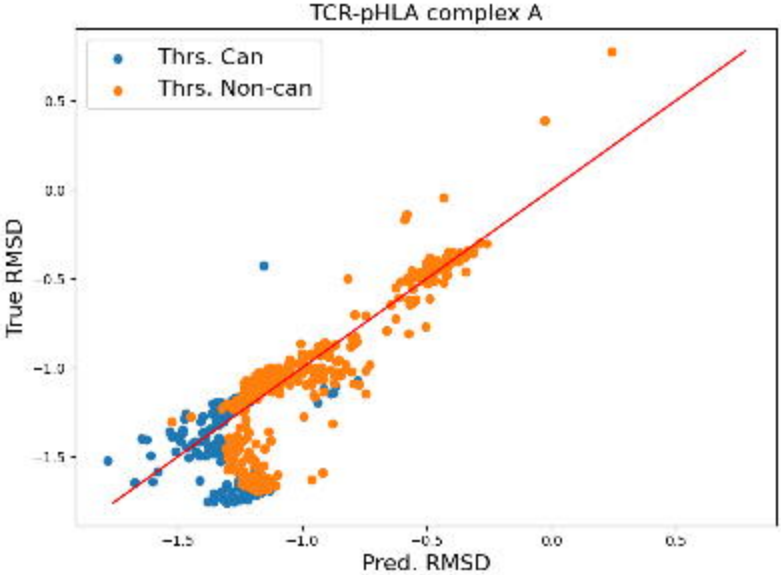
| | DoRIAT | DeepRank (naive-GNN) | DeepRank-GNN | DeepRank-GNN-esm | GNN-DOVE | EMLy™Dock |
|---|---|---|---|---|---|---|
| TCR-pHLA complex C | 0.24 | 0.16 | 0.02 | 0.03 | 0.04 | **0.29** |
| 1QRN | **0.54** | 0.51 | 0.02 | 0.02 | 0.03 | 0.42 |
| 2BNQ | **0.20** | 0.09 | 0.01 | 0.02 | 0.04 | 0.06 |
| 2VLJ | **0.52** | 0.21 | 0.05 | 0.02 | 0.20 | 0.06 |
| 3QFJ | 0.51 | **0.60** | 0.10 | 0.58 | 0.56 | 0.59 |
| 4MNQ | 0.54 | **0.57** | 0.38 | 0.03 | 0.03 | 0.03 |
| 5C07 | **0.37** | 0.26 | 0.03 | 0.03 | 0.02 | 0.03 |
| 5COA | **0.32** | 0.1 | 0.11 | 0.02 | 0.02 | 0.05 |
| 5C0B | **0.41** | 0.19 | 0.03 | 0.02 | 0.03 | 0.04 |
| 5C0C | **0.25** | 0.09 | 0.03 | 0.03 | 0.03 | 0.03 |
| 5EU6 | **0.41** | 0.17 | 0.19 | 0.1 | 0.09 | 0.03 |
| 5MEN | **0.64** | 0.10 | 0.43 | 0.02 | 0.02 | 0.28 |
| 5YXU | **0.64** | 0.62 | 0.06 | 0.58 | 0.5 | 0.54 |

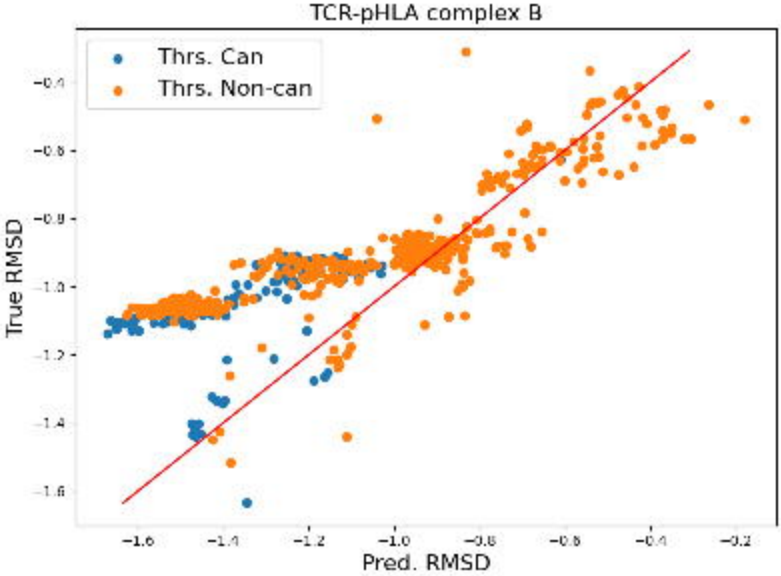

2PYE_ABCDE
Thrs. Can
Thrs. Non-can
True RMSD
Pred. RMSD
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0

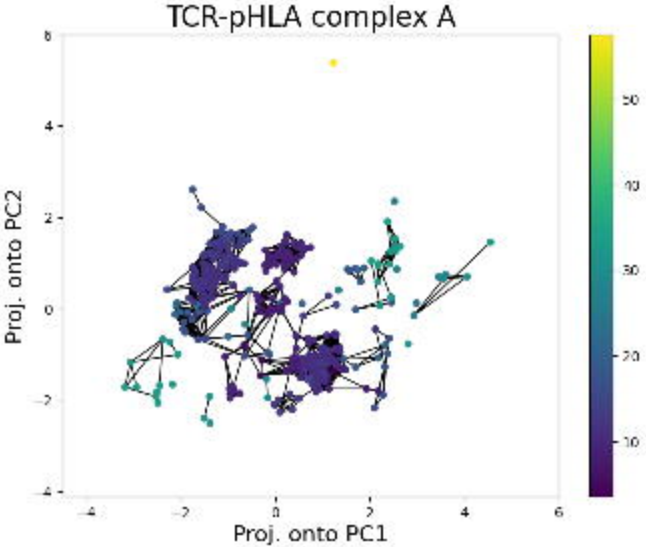

5ISZ_ABCDE
Thrs. Can
Thrs. Non-can
True RMSD
Pred. RMSD
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
−1.50
−1.75
−2.00

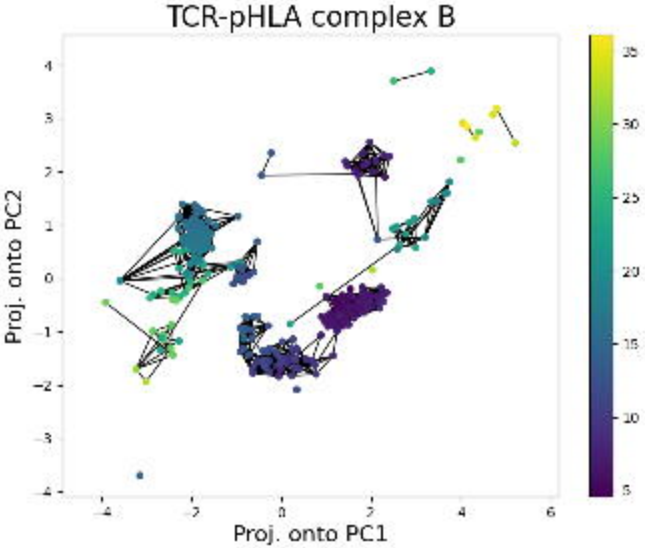

5NMG_ABCDE
Thrs. Can
Thrs. Non-can
True RMSD
Pred. RMSD
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5
−1.0
−0.5
0.0
0.5
1.0

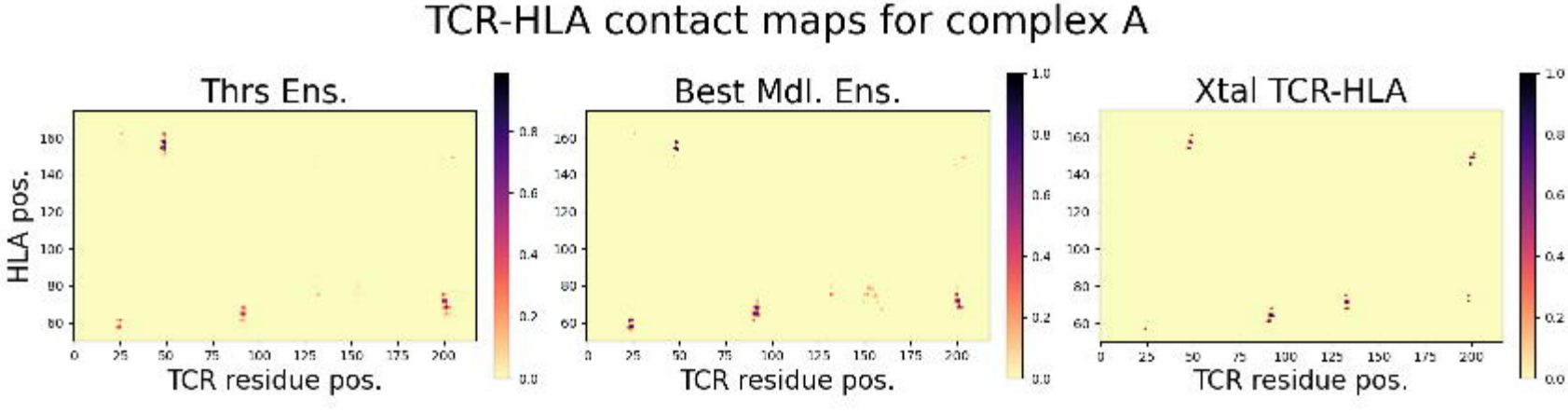

Cumulative Hitrate (%)
Chosen model within top k docking models
0
20
40
60
80
100
50
100
150
200
250
300
350
400
450
DeRIAT
DeepRank-naive GNN
DeepRank-GNN
DeepRank-GNN-esm
GNN DOVF
GMLy$^{TN}$Dock score

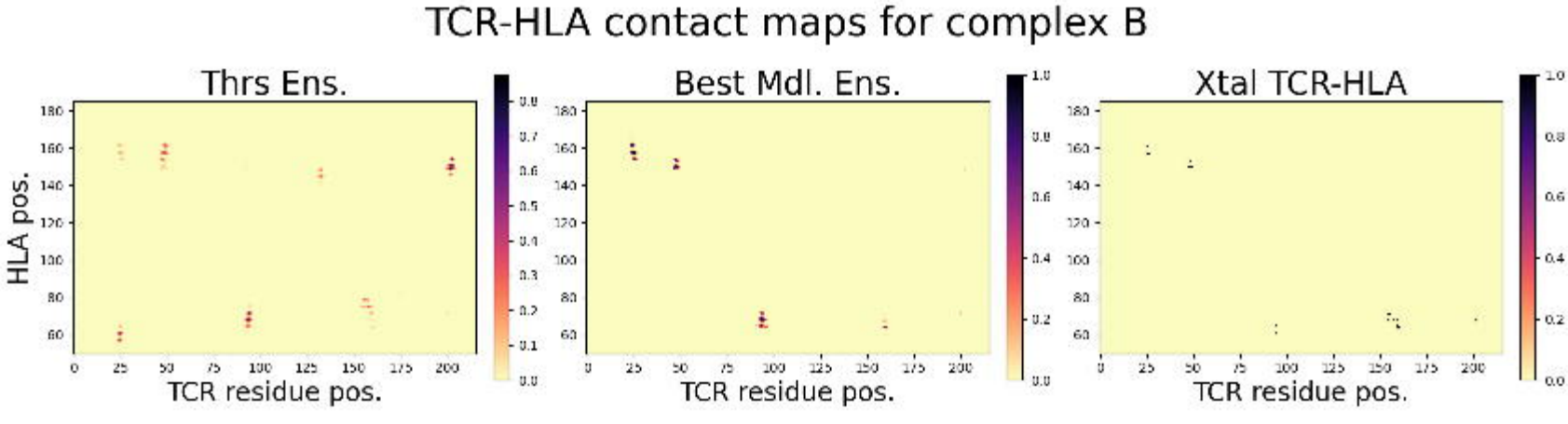

TCR-pHLA complex A
Thrs. Can
Thrs. Non-can
True RMSD
Pred. RMSD
0.5
0.0
−0.5
−1.0
−1.5
−1.5
−1.0
−0.5
0.0
0.5


TCR-pHLA complex B
Thrs. Can
Thrs. Non-can
True RMSD
Pred. RMSD
−0.4
−0.6
−0.8
−1.0
−1.2
−1.4
−1.6
−1.6
−1.4
−1.2
−1.0
−0.8
−0.6
−0.4
−0.2

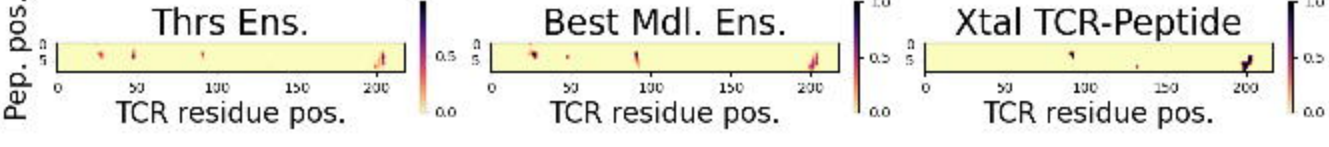

TCR-pHLA complex A
Proj. onto PC2
Proj. onto PC1
6
4
2
0
−2
−4
−4
−2
0
2
4
6
50
40
30
20
10


TCR-pHLA complex B
Proj. onto PC2
Proj. onto PC1
4
3
2
1
0
−1
−2
−3
−4
−4
−2
0
2
4
6
35
30
25
20
15
10
5

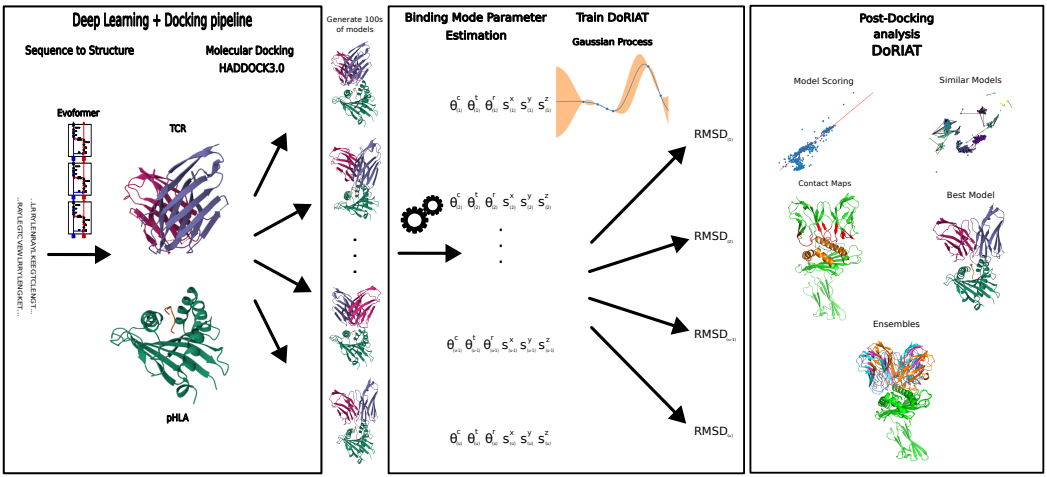

TCR-HLA contact maps for complex A
Thrs Ens.
Best Mdl. Ens.
Xtal TCR-HLA
HLA pos.
TCR residue pos.
TCR residue pos.
TCR residue pos.

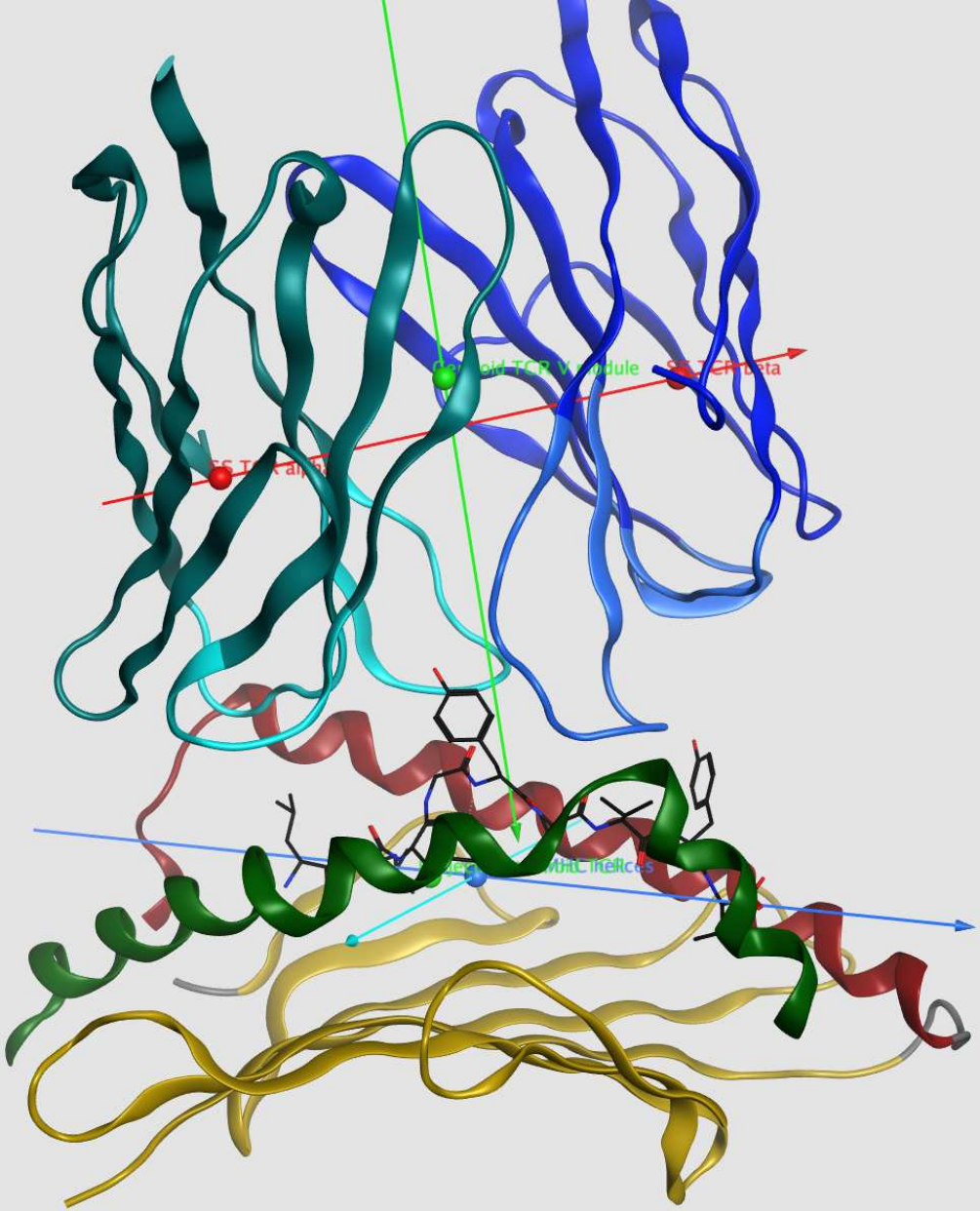

TCR-HLA contact maps for complex B
Thrs Ens.
Best Mdl. Ens.
Xtal TCR-HLA
HLA pos.
TCR residue pos.
TCR residue pos.
TCR residue pos.

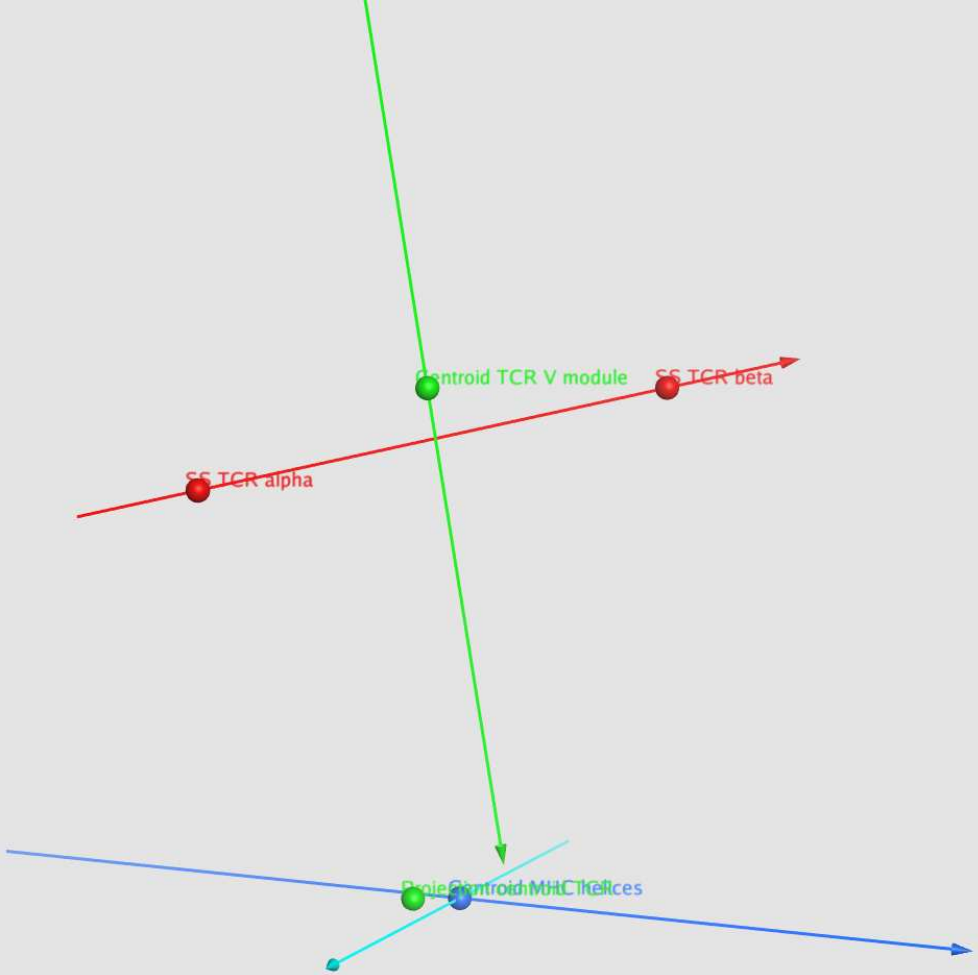
# TCR-Peptide contact maps for complex A

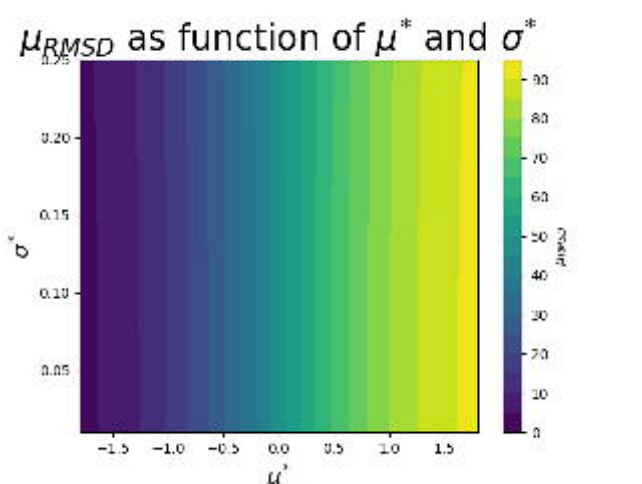


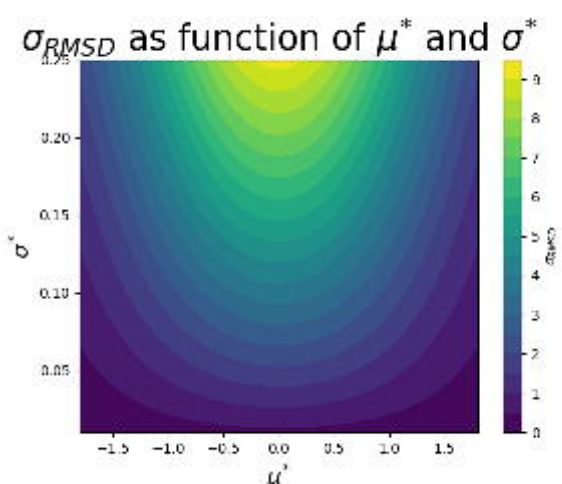
# TCR-Peptide contact maps for complex B

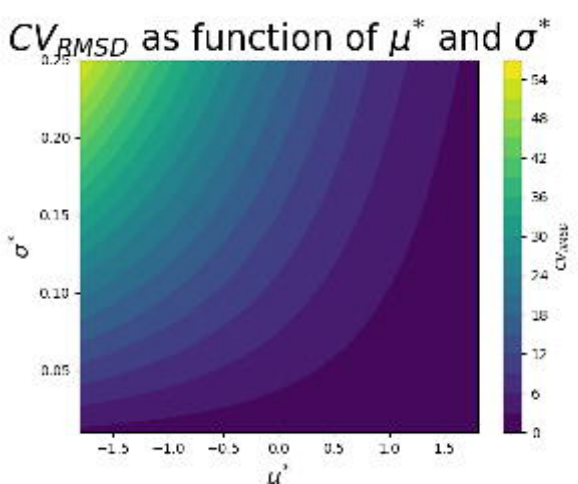

Deep Learning + Docking pipeline
Sequence to Structure
Molecular Docking
HADDOCK3.0
Evoformer
TCR
pHLA
Generate 100s of models
Binding Mode Parameter Estimation
Train DoRIAT
Gaussian Process
RMSD
Post-Docking analysis
DoRIAT
Model Scoring
Similar Models
Contact Maps
Best Model
Ensembles


Centroid TCR V module
SS TCR beta
SS TCR alpha
Mhc helices


Centroid TCR V module
SS TCR beta
SS TCR alpha
Proje
Centroid MHC helices


$\mu_{RMSD}$ as function of $\mu^*$ and $\sigma^*$
0.25
0.20
0.15
0.10
0.05
$\sigma^*$
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
$\mu^*$
90
80
70
60
50
40
30
20
10
0
$\mu_{RMSD}$


$\sigma_{RMSD}$ as function of $\mu^*$ and $\sigma^*$
0.25
0.20
0.15
0.10
0.05
$\sigma^*$
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
$\mu^*$
9
8
7
6
5
4
3
2
1
0
$\sigma_{RMSD}$


$CV_{RMSD}$ as function of $\mu^*$ and $\sigma^*$
0.25
0.20
0.15
0.10
0.05
$\sigma^*$
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
$\mu^*$
54
48
42
36
30
24
18
12
6
0
$CV_{RMSD}$