

DeepLigand: accurate prediction of MHC class I ligands using peptide embedding

Haoyang Zeng^{1,2} and David K. Gifford^{1,2,*}

¹Computer Science and Artificial Intelligence Laboratory and ²Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

*To whom correspondence should be addressed.

Abstract

Motivation: The computational modeling of peptide display by class I major histocompatibility complexes (MHCs) is essential for peptide-based therapeutics design. Existing computational methods for peptide-display focus on modeling the peptide-MHC-binding affinity. However, such models are not able to characterize the sequence features for the other cellular processes in the peptide display pathway that determines MHC ligand selection.

Results: We introduce a semi-supervised model, DeepLigand that outperforms the state-of-the-art models in MHC Class I ligand prediction. DeepLigand combines a peptide language model and peptide binding affinity prediction to score MHC class I peptide presentation. The peptide language model characterizes sequence features that correspond to secondary factors in MHC ligand selection other than binding affinity. The peptide embedding is learned by pre-training on natural ligands, and can discriminate between ligands and non-ligands in the absence of binding affinity prediction. Although conventional affinity-based models fail to classify peptides with moderate affinities, DeepLigand discriminates ligands from non-ligands with consistently high accuracy.

Availability and implementation: We make DeepLigand available at <https://github.com/gifford-lab/DeepLigand>.

Contact: gifford@mit.edu

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

The major histocompatibility complex (MHC) Class I molecules play a central role in the acquired immune system of vertebrates by presenting intracellular peptides on the cell surface for recognition by T cells (Castellino *et al.*, 1997; Janeway Jr *et al.*, 2001). The choice of which peptides are displayed is in part controlled by an individual's MHC genotype. MHC genes are highly polymorphic and thus peptide presentation varies from individual to individual (Jin and Wang, 2003; Williams, 2001). Modeling the individual-specific preference of peptide presentation by the MHC is important for designing efficient peptide-based immuno-therapeutics for individuals (Kreiter *et al.*, 2015; Ott *et al.*, 2017; Verdegaal *et al.*, 2016).

Class I MHCs present peptides derived from cytosolic proteins that are degraded by a proteasome. The degraded peptides are internalized by the transporter associated with antigen processing (TAP) channel in the endoplasmic reticulum to be potentially associated with class I MHC molecules. The formation of a MHC-peptide complex thus depends on multiple factors, including the proteasome cleavage preference of a peptide (Nielsen *et al.*, 2005), TAP

transport efficiency (Peters *et al.*, 2003) and the MHC-peptide binding affinity (Nielsen and Andreatta, 2016).

Most existing computational methods (Bhattacharya *et al.*, 2017; Han and Kim, 2017; Lundegaard *et al.*, 2008; Nielsen and Andreatta, 2016) for peptide presentation focus on modeling MHC-peptide binding affinity based on curated affinity datasets from the immunology epitope database (IEDB). Although binding affinity is essential for the formation of peptide-MHC complex, models trained only on peptide-MHC affinity are unable to characterize other factors that determine peptide display. Recent advances in mass spectrometry technology have enabled the high-throughput experimental identification of displayed peptides. Two recent computational models incorporate natural ligand datasets from mass spectrometry data in their models. MHCflurry (O'Donnell *et al.*, 2018) only considers the natural ligands as peptides that also have high binding affinity to the MHCs, and thus it remains an affinity model that doesn't consider other peptide features that influence peptide presentation. NetMHCpan4.0 (Jurtz *et al.*, 2017) is a modified affinity model that includes an additional output node for training on natural ligand observations from mass spectrometry data.

However, it assumes that affinity and uncorrelated display selection processes share the same sequence features.

Here we introduce DeepLigand, a semi-supervised approach to predicting natural MHC ligands with improved accuracy. We explicitly frame the identity of displayed ligands as a function of (i) MHC-peptide binding affinity and (ii) a vector embedding of the peptide that characterizes sequence patterns potentially associated with footprints of proteasome cleavage, TAP transport efficiency, and motifs relevant to other processes. For model element (i) we model MHC-binding affinity by a deep residual network that predicts affinity from the amino acid sequence of the MHC and peptide. For model element (ii) we learn an embedding of peptide sequences by performing unsupervised learning on the natural ligands of all MHCs. We show that DeepLigand improves the performance in natural MHC ligand prediction comparing to existing published methods as well as similar models that don't use the contextualized embedding of peptides. DeepLigand can identify MHC ligands from candidate peptides with moderate MHC-binding affinities, a challenging task where conventional affinity-based models struggle. Moreover, we demonstrate that the peptide embedding alone is highly predictive of natural ligands, indicating that meaningful patterns associated with ligand selection are captured in the embedding.

2 Materials and methods

2.1 Datasets

We used the MHC class I dataset curated by O'Donnell *et al.* (2018). This dataset consists of 525 672 binding affinity and mass spectrometry measurements collected from IEDB, Abelin *et al.* (2017) and Kim *et al.* (2014). This dataset also contains 2 541 370 non-ligand sequences (decoys) sampled from the protein-coding transcripts that also contained the mass spectrometry-identified peptides (hits) based on protein sequences in the UCSC hg19 proteome and transcript quantification from RNA sequencing of the relevant cell line (B721.221). As described in Abelin *et al.*, for an allele with n hits, $100n$ decoys of length 8–15 were sampled, weighting transcripts by the number of hits.

Every peptide training example is labeled by binding affinity and presence/absence in mass spectrometry data. Some of the examples in the binding affinity and mass spectrometry datasets have only qualitative affinity measurements, represented as positive, positive-high, positive-intermediate, positive-low or negative. For such examples, O'Donnell *et al.* assigned a quantitative value and a relation of '>' or '<' for each class of qualitative labels to denote the corresponding range of affinity. Examples with quantitative affinity were assigned a relation of '='. We further assigned an 'unknown' affinity to the non-ligand peptides. As is standard practice in the peptide-display literature (Jurtz *et al.*, 2017; Nielsen and Andreatta, 2016; O'Donnell *et al.*, 2018), we normalized the original affinity measurements in nano-molar (nM) by capping them between 1 and 50 000 nM and transforming with $1 - \log(\text{nM})/\log(50\,000)$ such that the resulting normalized affinities fall between zero and one. Aside from binding affinity, we also created a ligand label for each example. We labeled the ligands identified by mass spectrometry as positive, the non-ligand peptides as negative, and the other peptides as 'unknown'. We further removed MHC alleles for which the pseudo-sequence of MHC amino acids in contact with the peptide is not available from the literature (Jurtz *et al.*, 2017). The final dataset consists of 3 052 388 examples covering 219 MHC Class I alleles (Supplementary Fig. S1, Supplementary Table S2).

Consistent with previous work (Jensen *et al.*, 2018; Nielsen and Andreatta, 2016) we constructed a 5-fold cross-validation (CV) benchmark to avoid over-estimation of model performance. The benchmark minimizes the 8-mer overlap between different folds using a Hobohm *et al.* (1992) inspired algorithm described in Nielsen *et al.* (2007). The performance on a data point was evaluated by the prediction made from a model trained on the four CV-folds that don't contain the data point.

2.2 Model structure

DeepLigand consists of two modules, a binding affinity prediction module and a peptide embedding module (Fig. 1). The peptide embedding module embeds each peptide into a vector representation. Given a pair of MHC and peptide, we make the assumption that the measured binding affinity is the sum of the true affinity and a Gaussian noise with zero mean, and thus the measured affinity follows a Gaussian distribution with a mean at the true affinity. The affinity prediction module predicts the mean and variance of the Gaussian distribution which respectively characterizes the strength of affinity and the estimated level of observational noise associated with the input. Similar techniques have been used to characterize the observational noise in various types of data (aleatoric uncertainty) (Kendall and Gal, 2017; Lakshminarayanan *et al.*, 2017). Using the predicted affinity mean and variance as well as the peptide embedding as input features, we used one fully connected layer with sigmoid activation to predict whether the input peptide is a natural ligand of the input MHC. The peptide embedding module was trained on the natural ligands in the training set before it was used to generate peptide embedding for all training examples. Then the affinity module and the combining layer were jointly trained with the training loss calculated as the sum of the cross-entropy loss from ligand prediction and the log probability of the observed affinity under the predicted Gaussian affinity distribution. For both types of loss, examples with a 'unknown' label were not included in the calculation. As was adopted in O'Donnell *et al.*, for examples with an inequality relation for their affinity measurement, the loss was only included when the inequality is violated.

2.2.1 Binding affinity prediction module

The binding affinity prediction module is a deep residual network that takes as input the amino sequence of a peptide and the pseudo-sequence of a MHC molecule. As defined in the previous literature (Jurtz *et al.*, 2017), the pseudo-sequence of a MHC allele is the amino acid residues in 34 polymorphic positions where the residues are within 4.0 Å of the peptide in the structure of one or more of major MHC alleles. We represent each amino acid sequence by a 40-dimension feature vector concatenated from two encodings: a 20-dimension one-hot encoding and the 20-dimension BLOSUM50 (Nielsen *et al.*, 2003) matrix scores vector that depicts the evolutionary similarities between amino acids. All peptides are padded to 40 amino acids. For the padding amino acid, an all-zero vector was used for the one-hot encoding and a vector full of the lowest substitution score was used for the evolutionary encoding. This featureization leads to a 40×34 feature matrix for each MHC and a 40×40 feature matrix for each peptide. The MHC feature matrix (40×34) was then reshaped into a 1360×1 vector and appended to the peptide matrix (40×40) along its first dimension, resulting in a final feature matrix of 1400×40 for each MHC-peptide pair.

The binding affinity residual network consists of one initial convolutional block, five residual blocks and two branches of fully connected layers. Each residual block has two convolutional layers that

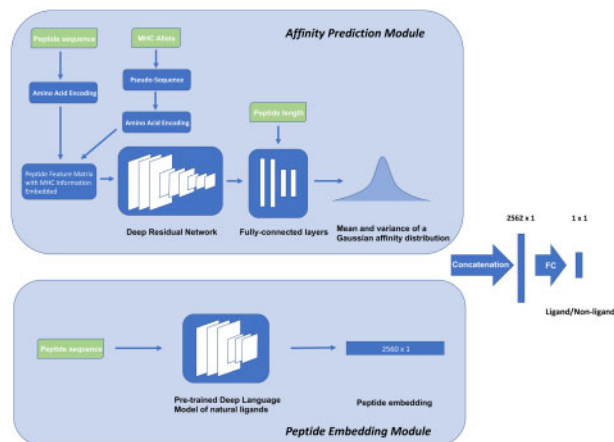


Fig. 1. Schematics of DeepLigand. DeepLigand consists of a binding affinity prediction module and a peptide embedding module. The affinity prediction module takes as input a pair of MHC and peptide sequence and predicts the mean and variance their binding affinity using a deep residual network. The peptide embedding module is a deep language model (ELMo) that embeds each peptide into a vector representation. The outputs from the two modules are concatenated and provided as input to one fully connected (FC) layer with sigmoid activation to predict whether the input peptide is a natural ligand of the input MHC. As a pre-training step, the embedding module is trained on the natural ligands in the training set for a given CV split. Then the affinity prediction module and the combining layer is jointly trained on all training examples taking as input the MHC and peptide sequences as well as the peptide embeddings produced by the pre-trained embedding module

fit the residual between the input and the output. Each convolutional layers in the network has 256 convolutional kernels with a stride of 1 and a kernel size of 3. Batch normalization was used after each convolutional layer. The output of the last residual block was flattened and concatenated with the sigmoid-transformed peptide length L and $1 - L$ as the input to the last two branches of fully connected layers, respectively to predict a Gaussian mean and variance of the binding affinity between the input MHC and peptide. Each branch consists of one layer of 64 neurons and one layer of 1 neuron. ReLU was used throughout the network as non-linearity and Adam (Kingma and Ba, 2014) was used as the optimizer. For each CV split, a random 1/9 of the training set was held out as the validation set. All hyper-parameters, including training epochs and optimizer parameters, were chosen according to the performance on the validation set. Using the peptide sequence, MHC pseudo-sequence, and the peptide embedding produced by the pre-trained peptide embedding module as input, this module was jointly trained with the last layer of DeepLigand for 50 epochs with early stopping when no improvement on the validation loss was observed for 10 epochs.

2.2.2 Peptide embedding module

In the peptide embedding module, we used ELMo (Peters *et al.*, 2018) to learn a context-dependent embedding of amino acids. Considering each peptide sequence as a sentence and each amino acid as a word, ELMo trains a deep bi-directional language model on the corpus (all natural ligand sequences). The ELMo embedding of each word in a sentence is the average of the hidden states of all layers in the language model for that word, which takes into account both the word and the semantic context. The language model used in this work is reduced in size compared with Peters *et al.* to accommodate the smaller corpus in this study. The details of the model structure can be found in Supplementary Table S1. For a given CV

split, a ELMo model was trained on all the natural ligands in the training set. The resulting model was then used to embed each peptide into a 64×40 matrix where 64 is the embedding dimension and 40 is the maximum length of the peptides in our dataset. The embedding matrix was then reshaped into an embedding vector of size 2560 as the final output of this module for a given peptide.

2.3 Implementation of alternative learning strategies

We implemented the two alternative learning strategies Affinity-Only and Two-Task (Section 3) by modifying the model structure of DeepLigand. Affinity-Only was implemented as the same structure as the affinity prediction module in DeepLigand. For Two-Task, we changed DeepLigand's affinity prediction module structure by adding an additional set of two fully connected layers with a sigmoid output to predict whether the input peptide is a ligand of the input MHC. This new branch has the same structure as the original two fully connected layers that predicts affinity, and also shares the same input features produced by the previous residual blocks. Embedding-Only was implemented as the last layer of DeepLigand with only the peptide embedding as input features.

2.4 Comparison with NetMHCpan4.0 and MHCflurry

The standalone version of NetMHCpan4.0 was downloaded from <http://www.cbs.dtu.dk/services/NetMHCpan/>. MHCflurry version 1.2.2 was installed as instructed on <https://github.com/openvax/mhcfurry> and the pre-trained 'models_class1_trained_with_mass_spec' model (https://github.com/openvax/mhcfurry/releases/download/pre-1.2.1/models_class1_trained_with_mass_spec.20180228.tar.bz2) was downloaded by 'mhcfurry-downloads fetch models_class1_rained_with_mass_spec'. Both methods only support certain MHC alleles. Only test set examples for which both methods can process were used in the model evaluation.

2.5 Sufficient Input Subsets analysis

The Sufficient Input Subset (SIS) software was downloaded from https://github.com/google-research/google-research/tree/master/sufficient_input_subsets. To accommodate the computation efficiency of SIS, we randomly sampled 10 000 MHC ligands with 9 amino acids and performed SIS on each ligand sequence to identify the minimal subset of residuals for the Embedding-Only model to predict the input sequence as ligand with >95% probability. The first layer of the language model in Embedding-Only represents each amino acid as a vector. As suggested by Carter *et al.*, we used the average of the vector representations for all 20 amino acids to represent the 'mask' residual that SIS uses to iteratively replace the original residuals in search for the minimally sufficient subset. The resulting SISs were clustered into six non-noise groups using DBSCAN in Scikit-Learn (Pedregosa *et al.*, 2011) with pre-computed Levenshtein distance as metric and hyper-parameters $\text{eps} = 1$ and $\text{min_samples} = 50$. WebLogo (Crooks *et al.*, 2004; <http://weblogo.threplusone.com>) was used to aggregate the SISs in each cluster into a sequence logo.

3 Results

3.1 DeepLigand improves the state-of-the-art accuracy in MHC ligand prediction

We trained and evaluated DeepLigand on a curated dataset of peptide-MHC-binding affinity measurements, mass spectrometry measurements, and non-ligand peptides collected by O'Donnell *et al.* (Section 2). As is standard practice in the peptide-display literature

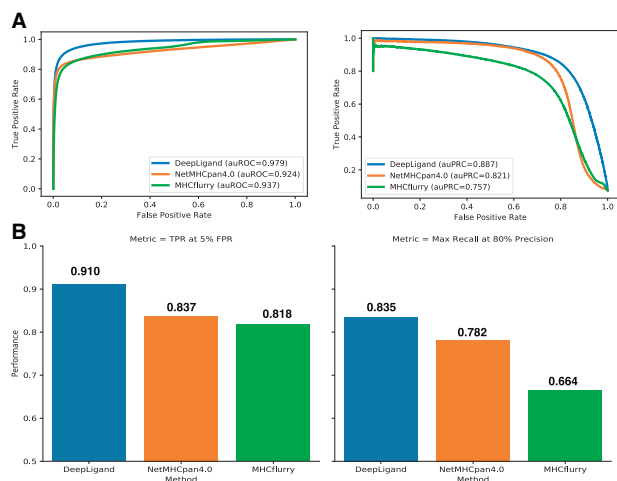


Fig. 2. Performance comparison with published state-of-the-art methods for peptide display. The evaluation was performed on test set examples that all three methods can process. **(A)** DeepLigand achieved higher auROC (left) and auPRC (right). **(B)** DeepLigand achieved a higher true positive rate at 5% false positive rate (left) and higher recall at 80% precision (right)

(Jensen *et al.*, 2018; Nielsen and Andreatta, 2016; Nielsen *et al.*, 2007), we constructed a 5-fold CV benchmark while minimizing the 8-mer overlap between different folds (Section 2).

We evaluated DeepLigand against NetMHCpan4.0 and MHCflurry. These two alternative models can only make predictions for select MHC alleles. We limited our test set samples to examples that all methods could process. We note that, while the test set is completely held out from DeepLigand, MHCflurry’s training set contains all the ligands in our test set, making the benchmark biased towards MHCflurry. NetMHCpan4.0 was trained on unreleased datasets and we do not know the overlap with our test set.

We found that DeepLigand outperforms previous state-of-the-art models at the task of predicting peptide presentation, despite the potential train-test set overlap advantages of previous methods (Fig. 2A). DeepLigand accurately predicts MHC ligands with an area under receiver operating characteristic curve (auROC) of 0.979 and an area under precision recall curve (auPRC) of 0.887, compared with an auROC of 0.937 (MHCflurry) and auPRC of 0.821 (NetMHCpan4.0) for the best previous state-of-the-art models. In therapeutic design, a high true positive rate at a low false positive rate and high recall with high precision are often more important metrics. DeepLigand outperforms the competing methods with a true positive rate of 91.0% at a false positive rate of 5% and 83.5% recall at 80% precision (Fig. 2B). The best alternative method (NetMHCpan4.0) produces a true positive rate of 83.7% at a false positive rate of 5%, and 78.2% recall at 80% precision.

3.2 DeepLigand outperforms alternative learning strategies for MHC ligand prediction

We next compared DeepLigand and alternative formulations of MHC ligand models with a controlled training set. MHCflurry only treats natural ligands identified by mass spectrometry as additional peptides with high binding affinity, and we refer to this strategy as *Affinity-Only*. NetMHCpan4.0 treats binding affinity and ligand presentation as two separate tasks that share all intermediate features, and we refer to this strategy as *Two-Task*. We implemented Affinity-Only and Two-Task by modifying the affinity prediction module in DeepLigand (Section 2). We trained and tested these models on the same datasets as DeepLigand, and thus excluded

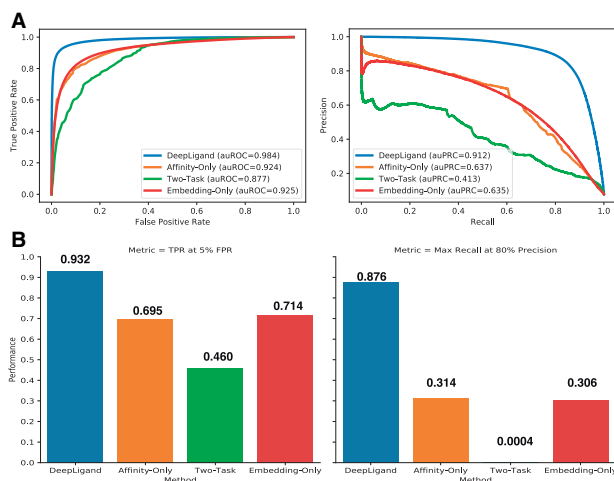


Fig. 3. Performance comparison with alternative learning strategies trained on the same dataset as DeepLigand. The evaluation was performed on all test set examples. **(A)** DeepLigand achieved higher auROC (left) and auPRC (right). **(B)** DeepLigand achieved a higher true positive rate at 5% false positive rate (left) and higher recall at 80% precision (right)

confounding factors such as training dataset mismatch to more precisely examine the power of different strategies for modeling MHC ligand. Note that unlike the comparison with published methods, here we were able to use the whole test set for evaluation.

We found that DeepLigand outperforms the alternative model formulations with an auROC of 0.984, an auPRC of 0.912, a true positive rate of 93.2% at a false positive rate of 5%, and 87.6% recall at 80% precision (Fig. 3). The best alternative model (Affinity-Only) achieved an auROC of 0.924, an auPRC of 0.637, a true positive rate of 69.5% at a false positive rate of 5% and 31.4% recall at 80% precision. Interestingly, our Two-Task implementation showed inferior performance compared with NetMHCpan4.0 which uses the same modeling strategy. Aside from the difference in the training set, this could arise from the fact that the non-ligand to ligand ratio in our benchmark dataset (100 \times) is much higher than the dataset that NetMHCpan4.0 was trained on (10 \times), which makes joint learning of both tasks more challenging.

Affinity-Only represents the modeling strategy of most peptide-display models used in therapeutic vaccine design to select for MHC-binding peptides [predicted affinity ≤ 500 nM; Ott *et al.* (2017); Łuksza *et al.* (2017)]. Therefore, we examined the performance of Affinity-Only and DeepLigand for MHC ligand prediction. We observed that the two formulations agree on low-affinity peptides. However, the predictions are much less consistent on peptides with medium to high predicted affinities (Fig. 4A). Evaluated on different subsets of peptides predicted to bind to the corresponding MHC, the predictive performance of Affinity-Only model degenerates fast towards random as we focused more on medium binders, while DeepLigand accurately discriminates ligands from non-ligands with consistently high auROCs and auPRCs above 0.9 (Fig. 4B).

3.3 Peptide embedding encapsulates sequences patterns predictive of natural ligands

We further examined the relative importance of DeepLigand’s peptide embedding in its MHC ligand prediction task. We removed the affinity prediction module in DeepLigand such that ligand prediction is made by a one-layer neural network using solely the peptide embedding as its features (Section 2). We refer to this strategy as *Embedding-Only*. We observed that the Embedding-Only model

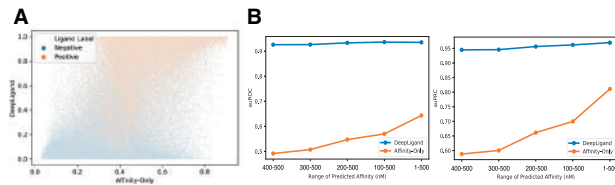


Fig. 4. Comparison between Affinity-Only and DeepLigand. **(A)** The scatter plot of affinity prediction from Affinity-Only (x-axis) and ligand prediction from DeepLigand (y-axis). Each point is a pair of peptide-MHC colored by whether the peptide is a ligand of the MHC as determined by mass spectrometry. **(B)** auROC (left) and auPRC (right) of DeepLigand and Affinity-Only when considering peptides with predicted affinities in different ranges. To obtain practically meaningful thresholds, affinities were converted to nM unit and affinity of lower nano-molar is stronger

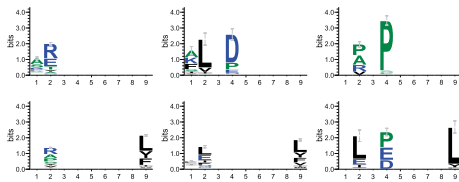


Fig. 5. The sequence logos of each of the six clusters of SISs identified from 10 000 random samples of MHC ligands of 9 amino acids. The x-axis denotes the position on the peptide and the y-axis denotes the information content

achieved high accuracy in MHC Class I ligand prediction (Fig. 3). As Embedding-Only takes no MHC information as input, the high predictive performance suggests that the embedding that DeepLigand learned from natural MHC ligands in an unsupervised fashion captures important sequence features shared across the ligands of different MHCs. These sequence features, which have been largely ignored by existing models, are highly discriminative of ligands and non-ligands even without knowledge about the affinity to the MHC molecule.

We applied Sufficient Input Subset (SIS) analysis (Carter *et al.*, 2018) to interpret the sequence features the Embedding-Only model has learned to identify MHC ligands. On 10 000 random samples of all MHC ligands of 9 amino acids in our dataset, we performed SIS to locate the minimal subset of residuals for the Embedding-Only model to predict a peptide as MHC ligand with a probability >95% (Section 2). Using DBSCAN as suggested by Carter *et al.*, the resulting SISs can be clustered into six groups. The sequence logos of these six groups suggest that the residuals at positions 1, 2, 4 and 9 of a peptide are most essential for the discrimination of ligands from non-ligands (Fig. 5A). Moreover, the motifs manifested by the SISs clusters characterize similar amino acid preferences as the known proteasome cleavage motif (Fig. 1 in Keşmir *et al.*, 2002; the upside down letters mean depletion), such as R/K/A at position 1, R/L/A/I/P at position 2, D/P/E at position 4 and L/Y/F/V at position 9.

4 Discussion

Existing computational efforts to model peptide presentation by MHC class I molecules have been focused on predicting the binding affinity between MHC molecules and peptides. However, additional factors determine which peptides in the proteome become natural MHC ligands, including the abundance of their transcripts, their proteasome cleavage motifs, and TAP-mediated peptide transport efficiency. Without considering these auxiliary factors, affinity-based

models could lead to false positives when designing MHC-presenting peptides in therapeutic design.

We presented a new semi-supervised model for the presentation of peptides by MHC Class I molecules. Our method DeepLigand explicitly frames the identity of MHC ligand as a function of binding affinity prediction and a sequence representation that characterizes auxiliary factors in ligand determination. We show that by employing a contextual word embedding algorithm trained on ligands in an unsupervised manner, we can embed peptides in a predictive vector representation. Even though only positive ligands were used during training, the resulting embedding alone is highly discriminative of natural ligands and non-ligands. Leveraging a published method for model interpretation, we find that the sequence patterns for which the embedding is highly indicative of ligands recapitulate the known proteasome cleavage motif, suggesting the sequence determinants of important factors in MHC ligand selection other than binding affinity have been learned. Combining the peptide embedding with binding affinity predicted from the MHC and peptide sequence, DeepLigand improved the accuracy in natural MHC ligand prediction compared previously published state-of-the-art.

We observe that widely used affinity-based models have a high false positive rate when prioritizing natural ligands from peptides that have medium binding affinities. By point of comparison, the ligand prediction from DeepLigand, which uses the peptide embedding obtained from unsupervised learning on natural ligands, can still accurately discriminate held-out ligands from non-ligands.

With DeepLigand's improved false positive rate, we envision DeepLigand can be used to design effective neo-antigen vaccine formulations. In neo-antigen vaccine design, cancer-specific peptides are prioritized for MHC presentation. Most of the cancer-specific peptides have low to medium binding affinity to the MHC molecules in the patient, and the number of administered peptides to be designed is highly limited (Ott *et al.*, 2017). Thus reducing the false positive rate will reduce the number of administered peptides that are not disease related.

Mass spectrometry is currently the most direct measure of natural MHC ligand presentation, and in this work we equate the read out from mass spectrometry with the collection of natural ligands. We note that certain biases (Klont *et al.*, 2018; Mahoney *et al.*, 2011) are known to exist in the sequences identified from mass spectrometry, which could limit the power of computational models trained on such datasets. With the future advance of biotechnology, we envision our approach can be further improved to deliver more accurate natural MHC ligand predictions.

Acknowledgements

We thank Brooke Huisman for helpful discussions. We also thank Brandon Carter for the guidance on running Sufficient Input Subsets analysis.

Funding

This work was supported by National Institute of Health [R01CA218094].

Conflict of Interest: D.K.G. is a founder of Think Therapeutics, a company that uses machine learning for therapeutic design.

References

- Abelin, J.G. *et al.* (2017) Mass spectrometry profiling of HLA-associated peptides in mono-allelic cells enables more accurate epitope prediction. *Immunity*, **46**, 315–326.
- Bhattacharya, R. *et al.* (2017) Evaluation of machine learning methods to predict peptide binding to MHC Class I proteins. *bioRxiv*, 154757.

- Carter, B. *et al.* (2018). What made you do this? understanding black-box decisions with sufficient input subsets. In *Proceedings of Machine Learning Research*, pp. 567–576.
- Castellino, F. *et al.* (1997) Antigen presentation by MHC class II molecules: invariant chain function, protein trafficking, and the molecular basis of diverse determinant capture. *Hum. Immunol.*, **54**, 159–169.
- Crooks, G.E. *et al.* (2004) Weblogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
- Han, Y. and Kim, D. (2017) Deep convolutional neural networks for pan-specific peptide-MHC class I binding prediction. *BMC Bioinformatics*, **18**, 585.
- Hobohm, U. *et al.* (1992) Selection of representative protein data sets. *Protein Sci.*, **1**, 409–417.
- Janeway, C.A. Jr. *et al.* (2001). The major histocompatibility complex and its functions. In *Immunobiology: The Immune System in Health and Disease*, 5th edn. Garland Science.
- Jensen, K.K. *et al.* (2018). Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology*.
- Jin, P. and Wang, E. (2003). Polymorphism in clinical immunology—From HLA typing to immunogenetic profiling. *J. Transl. Med.*, **1**, 8.
- Jurtz, V. *et al.* (2017). NetMHCpan-4.0: improved peptide—MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.*, [ji1700893](https://doi.org/10.1093/immk/ji1700893).
- Keşmir, C. *et al.* (2002) Prediction of proteasome cleavage motifs by neural networks. *Protein Eng.*, **15**, 287–296.
- Kendall, A. and Gal, Y. (2017) What uncertainties do we need in Bayesian deep learning for computer vision? In: Guyon, I. *et al.*, (eds) *Advances in Neural Information Processing Systems*. Curran Associates, Inc., pp. 5574–5584.
- Kim, Y. *et al.* (2014) Dataset size and composition impact the reliability of performance benchmarks for peptide-MHC binding predictions. *BMC Bioinformatics*, **15**, 241.
- Kingma, D.P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv: 1412.6980*.
- Klont, F. *et al.* (2018) Assessment of sample preparation bias in mass spectrometry-based proteomics. *Anal. Chem.*, **90**, 5405–5413.
- Kreiter, S. *et al.* (2015) Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature*, **520**, 692–696.
- Lakshminarayanan, B. *et al.* (2017) Simple and scalable predictive uncertainty estimation using deep ensembles. In: Guyon, I. *et al.* (eds) *Advances in Neural Information Processing Systems*. Curran Associates, Inc., pp. 6402–6413.
- Łuksza, M. *et al.* (2017) A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature*, **551**, 517.
- Lundegaard, C. *et al.* (2008) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res.*, **36**, W509–W512.
- Mahoney, D.W. *et al.* (2011) Relative quantification: characterization of bias, variability and fold changes in mass spectrometry data from iTRAQ-labeled peptides. *J. Proteome Res.*, **10**, 4325–4333.
- Nielsen, M. and Andreatta, M. (2016) NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med.*, **8**, 33.
- Nielsen, M. *et al.* (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.*, **12**, 1007–1017.
- Nielsen, M. *et al.* (2005) The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics*, **57**, 33–41.
- Nielsen, M. *et al.* (2007) NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS One*, **2**, e796.
- O'Donnell, T.J. *et al.* (2018) MHCflurry: open-source class I MHC binding affinity prediction. *Cell Syst.*, **7**, 129–132.
- Ott, P.A. *et al.* (2017) An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature*, **547**, 217–221.
- Pedregosa, F. *et al.* (2011) Scikit-learn: machine learning in python. *J. Mach. Learn. Res.*, **12**, 2825–2830.
- Peters, B. *et al.* (2003) Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J. Immunol.*, **171**, 1741–1749.
- Peters, M.E. *et al.* (2018) Deep contextualized word representations. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long Papers), pp. 2227–2237. Association for Computational Linguistics, New Orleans, Louisiana.
- Verdegaal, E.M.E. *et al.* (2016) Neoantigen landscape dynamics during human melanoma—T cell interactions. *Nature*, **536**, 91–95.
- Williams, T.M. (2001) Human leukocyte antigen gene polymorphism and the histocompatibility laboratory. *J. Mol. Diagn.*, **3**, 98–104.