

1 ConBoTNet: supervised contrastive learning enhances MHC-II peptide binding affinity prediction

3 Long-Chen Shen¹, Yan Liu², Zi Liu¹, Yumeng Zhang^{3,4,5}, Zhikang Wang^{3,4}, Yuming Guo⁶, Jamie
4 Rossjohn³, Jiangning Song^{3,4,*}, and Dong-Jun Yu^{1,*}

5 ¹School of Computer Science and Engineering, Nanjing University of Science and Technology, 200
6 Xiaolingwei, Nanjing, 210094, China;

7 ²Department of Computer Science, Yangzhou University, Yangzhou, 225100, China;

8 ³Monash Biomedicine Discovery Institute and Department of Biochemistry and Molecular Biology,
9 Monash University, Melbourne, VIC 3800, Australia;

10 ⁴Monash Data Futures Institute, Monash University, Melbourne, VIC 3800, Australia;

11 ⁵School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240,
12 China;

13 ⁶School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria,
14 Australia.

15

16

17 * To whom correspondence should be addressed.

18 Corresponding authors. Dong-Jun Yu, School of Computer Science and Engineering, Nanjing
19 University of Science and Technology, China. Email: njyudj@njust.edu.cn; Jiangning Song,
20 Biomedicine Discovery Institute and Department of Biochemistry and Molecular Biology, Monash
21 University, Melbourne, VIC 3800, Australia. Email: jiangning.song@monash.edu.

22

23

24 **Keywords:** MHC-II peptide binding; supervised contrastive learning; binding affinity prediction;
25 binding core identification

26

27

28 **Running Head:** supervised contrastive learning enhances MHC-II peptide binding affinity prediction

29 **Author Biography:**

30 **Long-Chen Shen** received his M.S. degree in software engineering from Nanjing University of
31 Science and Technology in 2021. He is currently a PhD candidate in the School of Computer Science
32 and Engineering, Nanjing University of Science and Technology and a member of the Pattern
33 Recognition and Bioinformatics Group. His current interests include pattern recognition, data mining,
34 and bioinformatics.

35 **Yan Liu** received his Ph.D. degree in computer science from Nanjing University of Science and
36 Technology in 2023. He is currently a young hundred distinguished professor with the School of
37 Information Engineering, Yangzhou University, China. His research interests include pattern
38 recognition, machine learning, and bioinformatics.

39 **Zi Liu** received his M.S. degree from Jingdezhen Ceramic University in 2015. From 2017 to 2019,
40 he pursued a joint Ph.D. program at the University of Michigan, USA. He is currently a Ph.D. student
41 at Nanjing University of Science and Technology. His principal research interests lie in bioinformatics,
42 machine learning, and pattern recognition.

43 **Yumeng Zhang** is a Master's student in the School of Life Sciences and Biotechnology at Shanghai
44 Jiao Tong University. He received his bachelor's degree in bioinformatics from Shanghai Jiao Tong
45 University. His research interests include machine learning, computational biology, and
46 bioinformatics.

47 **Zhikang Wang** is a PhD student in the Monash Biomedicine Discovery Institute and Department of
48 Biochemistry and Molecular Biology, Monash University, Melbourne, Australia. His research
49 interests include bioinformatics, machine learning and medical imaging.

50 **Yuming Guo** is a Professor of Global Environmental Health and Biostatistics & Head of the Monash
51 Climate, Air Quality Research (CARE) Unit. His research focuses on environmental epidemiology,
52 biostatistics, global environmental change, air pollution, climate change, urban design, residential
53 environment, remote sensing modeling and infectious disease modeling.

54 **Jamie Rossjohn** is a professor and group leader in the Monash Biomedicine Discovery Institute and
55 Department of Biochemistry and Molecular Biology, Monash University, Australia. His research is

56 focused on understanding the structural and biophysical basis of MHC restriction, T-cell receptor
57 (TCR) engagement and the structural correlates of T-cell signalling.

58 **Jiangning Song** is an Associate Professor and group leader in the Monash Biomedicine Discovery
59 Institute and Department of Biochemistry and Molecular Biology, Monash University, Melbourne,
60 Australia. He is also a member of the Monash Data Futures Institute. His research interests include
61 bioinformatics, computational biomedicine, machine learning, data mining, and pattern recognition.

62 **Dong-Jun Yu** received the PhD degree from Nanjing University of Science and Technology in 2003.
63 He is currently a full professor in the School of Computer Science and Engineering, Nanjing
64 University of Science and Technology. His research interests include pattern recognition, machine
65 learning and bioinformatics. He is a senior member of the China Computer Federation (CCF) and a
66 senior member of the China Association of Artificial Intelligence (CAAI).

67

68 **Abstract**

69 Accurate prediction of major histocompatibility complex (MHC)-peptide binding affinity could
70 provide essential insights into cellular immune responses and guide the discovery of neoantigens and
71 personalized immunotherapies. Nevertheless, the existing deep learning-based approaches for
72 predicting MHC-II peptide interactions fall short of satisfactory performance and offer restricted
73 model interpretability. In this study, we propose a novel deep neural network, termed ConBoTNet, to
74 address the above issues by introducing the designed supervised contrastive learning and bottleneck
75 transformer extractors. Specifically, the supervised contrastive learning pre-training enhances the
76 model's representative and generalizable capabilities on MHC-II peptides by pulling positive pairs
77 closer and pushing negative pairs further in the feature space, while the bottleneck transformer
78 module focuses on MHC-II peptide interactions to precisely identify binding cores and anchor
79 positions in an unsupervised manner. Extensive experiments on benchmark datasets under 5-fold
80 cross-validation, leave-one-molecule-out validation, independent testing, and binding core prediction
81 settings highlighted the superiority of our proposed ConBoTNet over current state-of-the-art methods.
82 Data distribution analysis in the latent feature space demonstrated that supervised contrastive learning
83 can aggregate MHC-II-peptide samples with similar affinity labels and learn common features of
84 similar affinity. Additionally, we interpreted the trained neural network by associating the attention
85 weights with peptides and innovatively find both well-established and potential peptide motifs. This
86 work not only introduces an innovative tool for accurately predicting MHC-II peptide affinity, but
87 also provides new insights into a new paradigm for modeling essential biological interactions,
88 advancing data-driven discovery in biomedicine.

89

90 INTRODUCTION

91 Major histocompatibility complex (MHC) molecules are a class of cell surface glycoproteins that play
92 an essential role in the immune system and are mainly divided into two types: MHC class I (MHC-I)
93 and MHC class II (MHC-II) [1, 2]. Among them, MHC-II molecules can recognize exogenous
94 antigens derived from pathogenic microorganisms or self-tissues to form MHC-peptide complexes,
95 and then present them on the surface of antigen-presenting cells (APC) to CD4⁺ T cell receptors
96 (TCR), thereby activating specific T cell immune responses and eliminating pathogens or damaged
97 tissues [3]. Accurate identification of MHC binding peptides is thus crucial for not only identifying
98 neoepitopes that effectively trigger immune response targeting cancer or infected cells but also
99 providing key insights into vaccine design [4], immunotherapy [5], and research on immune-related
100 diseases [6]. In this context, computational methods can assist in identifying potential peptides for
101 effective immune response, enhancing vaccine safety and efficacy, and guiding treatments for
102 immune-related diseases such as autoimmune disorders, allergies, and cancer [1, 7, 8].

103 The MHC-II molecule consists of two unique polypeptide chains, i.e., α chain and β chain. Each
104 chain has an external domain that together forms a peptide-binding groove to bind an antigenic
105 peptide [9]. MHC-II molecules in humans are encoded by various Human Leukocyte Antigen (HLA)
106 genes like HLA-DR, HLA-DQ and HLA-DP. Each MHC-II gene possesses numerous alleles,
107 resulting in substantial genetic diversity within the population. The diversity of MHC-II molecules
108 confers several benefits: (1) Improved antigen presentation, enabling a broader range of antigenic
109 peptides to be presented and thereby enhancing the capacity of the immune system to detect and
110 eradicate pathogens; (2) Facilitation of population adaptability, assisting organisms in adjusting to
111 diverse environmental and pathogenic pressures; (3) MHC-II molecules typically present peptides
112 derived from exogenous antigens. Unlike MHC-I, the open groove of MHC-II allows for
113 hypervariable peptide lengths, usually between 10-30 amino acids, with a typical range of 13-17
114 residues [10]. In essence, the polypeptides' ends extend past the binding groove, with only some key
115 amino acid residues (i.e., anchor residues) tightly coupled with MHC-II molecules, ensuring the
116 stability of protein-peptide interaction [11]. This characteristic allows MHC-II molecules to
117 effectively present exogenous antigens to CD4⁺ T cells, thus regulating immune responses. However,
118 experimental characterization of the binding specificity with numerous MHC-II molecules is time-

119 consuming and laborious. In this regard, computational methods can be employed as effective
120 alternatives to quantify peptide-MHC-II molecule binding affinity. Recent studies indicate the
121 growing importance of such computational methods and tools in the fields of vaccine design and
122 immunotherapy [12-14].

123 MHC-II peptide binding prediction methodologies mainly comprise allele-specific and pan-
124 specific methods. While allele-specific methods are restricted to predicting binding affinity for MHC-
125 II molecules found in the training set, pan-specific methods can predict unseen MHC-II molecules
126 with known protein sequences [15]. Pan-specific methods have broader applications and encompass
127 allele-specific methods. The classic allele-specific method, NetMHCII [15], trained separate artificial
128 neural networks (ANNs) for each MHC-II molecule to predict peptide-binding affinity. Given the
129 scarcity of experimental data for most MHC-II molecules and the limited scope of allele-specific
130 methods, pan-specific approaches emerged for MHC-II peptide binding prediction [16, 17].
131 NetMHCIIpan-3.2 [15] was the first pan-specific method to predict binding affinity using ANN. It
132 combines MHC-II pseudo-sequence information and integrates 40 networks with different numbers
133 of hidden neurons. Zeng *et al.* proposed a multi-model ensemble method PUFFIN [16] that quantified
134 prediction uncertainty and prioritized peptides with “binding likelihood” to improve the accuracy of
135 high-affinity peptide selection. MHCAttnNet [18] incorporated variable-length peptide sequences
136 and the MHC allele via attention mechanism and Bi-LSTM encoder, thus improving MHC-I and II
137 peptide binding predictions. Similarly, DeepSeqPanII [19], a recurrent neural network model with
138 attention mechanism, also predicted peptide-HLA class II binding. Recently, You *et al.* [17] proposed
139 a binding core-aware interaction model to enhance MHC-II peptide binding affinity prediction.
140 Despite impressive advances in prediction performance, current methods have some drawbacks: (1)
141 the large diversity of MHC-II molecules and scarcity of binding data often lead to low prediction
142 accuracy with limited training data; (2) while some algorithms aim to enhance model interpretability
143 by analyzing attention modules, there is still much room for improvement; (3) most methods
144 singularly characterize MHC-II molecules and peptide sequences, which may lead to insufficient
145 mining of sequence information; (4) Most algorithms currently ensemble multiple models of similar
146 structure to boost model performance. However, it will save computing resources for training and
147 prediction if comparable or superior performance can be achieved with fewer integrations or a single
148 model.

149 Here, we present ConBoTNet (Contrastive Bottleneck Transformer Network), a new deep-
150 learning framework that employs powerful supervised contrastive learning with the bottleneck
151 transformer extractor to enable accurate prediction of MHC-II peptide binding affinity. In the
152 supervised contrastive learning pre-training phase, the continuous values of binding affinity are
153 divided into ordered categories, and then the supervised contrastive loss function is applied to
154 minimize the normalized embeddings within each class and maximizes the distances between
155 different classes. Moreover, the bottleneck transformer modules can detect the binding core and
156 precise anchor positions from MHC-II molecules and peptide interaction features due to their
157 robustness in feature extraction. Performance evaluations, including 5-fold cross-validation (5-fold
158 CV), leave-one-molecule-out (LOMO) validation, and independent testing on benchmark datasets,
159 reveal that our model outperforms five current state-of-the-art (SOTA) methods, exhibiting rapid
160 training convergence and impressive generalization capacity. The source code is accessible at
161 <https://github.com/shenlongchen/conbotnet>.

162

163 MATERIALS AND METHODS

164 Benchmark datasets

165 In this study, three public benchmark datasets (BD2016, BC2015, BD2023) were used as benchmark
166 datasets to evaluate and compare the performance with other competing methods. The sources and
167 construction process of these three datasets are introduced in this section.

168 **BD2016:** Derived from NetMHCIIpan-3.2 [15], this dataset consists of MHC-II peptide binding
169 affinity data originally extracted from the IEDB database [20] (<https://www.iedb.org/>) in 2016.
170 Next, the binding values (IC_{50}) of the experimental data were normalized to [0,1] by $1 - \log(IC_{50} nM)/\log(50,000)$. BD2016 contains 134,281 MHC-II peptide affinity data for 36 HLA-
171 DR, 27 HLA-DQ, 9 HLA-DP, and 8 H-2 mouse molecules. It is considered the gold standard dataset
172 by previous affinity prediction methods, offering widely accepted 5-fold CV data partitioning by
173 grouping the peptides with common motifs into the same fold [16, 17, 19]. BD2016 is publicly
174 available at <https://services.healthtech.dtu.dk/services/NetMHCIIpan-3.2/>.

176 **BC2015:** This benchmark dataset contained 51 MHC-II peptide complexes with crystal structures

177 and has been used by several MHC-II peptide affinity prediction algorithms for unsupervised binding
178 core evaluations. It is also publicly available at
179 <https://services.healthtech.dtu.dk/services/NetMHCIIpan-3.2/>.
180 **BD2023**: an up-to-date independent test set curated from an automatically updated MHC-II binding
181 prediction platform (http://tools.iedb.org/auto_bench/mhcii/weekly/) [21]. Specifically, we selected
182 the newly added MHC-II peptide binding data generated by the platform from 2022-04-01 to 2023-
183 04-21, removed duplicates from the training set BD2016, and then filtered out MHC-II molecules
184 based on the number of corresponding peptide samples no less than 28 (This threshold ensures a rich
185 assortment of MHC-II categories, and also the rationality for calculating the predictive performance
186 index of each MHC-II molecule). The dataset contained samples of two measurement types, i.e., IC_{50}
187 and binary (binding or decoy) labels. Performance evaluation metrics used by the experiments are
188 provided in **Supplementary Text S1**.

189

190 Overview

191 The deep-learning framework of ConBoTNet is illustrated in **Figure 1**, which mainly includes the
192 following modules: (1) sequence encoding for peptides and MHC-II molecules; (2) backbone encoder
193 network for feature extraction; (3) projection network used for contrastive learning; (4) binding core
194 and affinity predictor. The core contributions of ConBoTNet are as follows: (i) Combining supervised
195 contrastive learning with a deep transformer-based network, we could effectively learn the latent
196 representations of the MHC-II peptide binding complex. (ii) Then, we can smoothly adapt the pre-
197 trained model to various downstream tasks, including binding affinity prediction and binding core
198 identification.

199

200 Sequence encoding

201 The input to ConBoTNet consists of peptide and MHC-II molecule sequence pairs. In particular, the
202 MHC-II molecule sequence is represented by a pseudo-sequence comprising 34 amino acids in
203 contact with the peptide-binding core (distance within 4Å) [22]. The first 15 amino acids of these are
204 from the α -chain while the rest are from the β -chain, following the strategy used in most previous
205 studies [15-17, 23].

206 We define P as the input peptide sequence with a length of L and Q as the MHC-II pseudo-
207 sequence with a length of 34, respectively. To capture the hidden semantic information and ensure
208 the complete representation of the input information, we encode both the peptide sequence and the
209 MHC-II pseudo-sequence using the embedding layer and one-hot encoding.

210 For the embedding layer, let d be the dimension of amino acid embeddings, $\mathbf{X}_e \in \mathbb{R}^{L \times d}$, the
211 output of the embedding layer for peptide sequence P , and $\mathbf{Y}_e \in \mathbb{R}^{34 \times d}$, the output of the embedding
212 layer for the MHC-II pseudo-sequence Q , are given as follows:

213
$$\mathbf{X}_e = (\mathbf{x}_e^1, \mathbf{x}_e^2, \dots, \mathbf{x}_e^L)^T, \mathbf{Y}_e = (\mathbf{y}_e^1, \mathbf{y}_e^2, \dots, \mathbf{y}_e^{34})^T \quad (1)$$

214 where $\mathbf{x}_e^i \in \mathbb{R}^d$ and $\mathbf{y}_e^j \in \mathbb{R}^d$ are the feature vectors of the i th amino acid of P and j th amino acid of
215 Q , respectively.

216 In terms of one-hot encoding, $d'=20$ is the dimension of one-hot encoding, representing 20
217 common amino acids. $\mathbf{X}_o \in \mathbb{R}^{L \times d'}$, the one-hot encoding for peptide sequence P , and $\mathbf{Y}_o \in \mathbb{R}^{34 \times d'}$,
218 the one-hot encoding for MHC-II pseudo-sequence Q , are given as follows:

219
$$\mathbf{X}_o = (\mathbf{x}_o^1, \mathbf{x}_o^2, \dots, \mathbf{x}_o^L)^T, \mathbf{Y}_o = (\mathbf{y}_o^1, \mathbf{y}_o^2, \dots, \mathbf{y}_o^{34})^T \quad (2)$$

220 where $\mathbf{x}_o^i \in \mathbb{R}^{d'}$ and $\mathbf{y}_o^j \in \mathbb{R}^{d'}$ are the one-hot feature vectors of the i th amino acid of P and j th amino
221 acid of Q , respectively. A value of 1 was assigned to the corresponding amino acid position in the
222 input sequences and 0 elsewhere.

223

224 Backbone encoder network

225 Next, we will briefly introduce the architecture of ConBoTNet in combination with **Figure 1b**. Firstly,
226 we used two feature encoding methods, learnable embedding, and one-hot encoding to characterize
227 the peptide sequence and the MHC-II molecular pseudo-sequence, respectively. Subsequently, we
228 built the interaction features through the interaction module proposed by DeepMHCII [17]. It is
229 noteworthy that we added 64 1×3 convolution filters after one-hot encoded features for peptide and
230 MHC-II, respectively. We then combined the interaction features from different encoding methods as
231 the final peptide MHC-II molecule interaction features. Next, these were sequentially passed through
232 two bottleneck transformers, several one-dimensional ResNet blocks, and an avg-pooling layer.
233 We refer to the above pipeline as the backbone encoder network, denoted as $Enc(\cdot)$, which is

234 expressed by the following formula:

$$235 \quad \mathbf{rep} = Enc(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{D_E} \quad (3)$$

236 We define the number of samples in a batch as N . \mathbf{x}, \mathbf{y} represent pairs of peptide and MHC-II
237 molecules. The encoder network map \mathbf{x}, \mathbf{y} to a representation vector \mathbf{rep} , which is normalized to the
238 unit hypersphere in \mathbb{R}^{D_E} ($D_E = 256$, according to the complexity of features).

239 **The details of Bottleneck transformers.** The current state-of-the-art methods for predicting MHC-II
240 peptide binding utilize shallow or deep convolutional networks, such as PUFFIN [16], DeepMHCII
241 [17]. However, we propose that incorporating the transformer module can further enhance the
242 predictions. Transformers capture long-term dependencies in MHC-II peptide binding features and
243 enhance context awareness through multi-head self-attention, enabling them to learn the universal
244 pattern and contextual information of MHC-II peptide binding. Inspired by the bottleneck transformer
245 (BoTNet)[24] architecture and its outstanding performance in computer vision tasks, we integrated it
246 with the domain knowledge of affinity prediction, and improved its Multi-head Self-Attention
247 (MHSA) layer to capture the contextual and global information related to MHC-II peptide interaction.

248 We illustrated an MHSA layer in **Figure 1c**. The multi-head self-attention layer can be
249 summarized by the following equation:

$$250 \quad \begin{aligned} MultiHead(X) &= Concat(head_1, \dots, head_h) \\ head_i &= Attention(XW_i^Q, XW_i^K, XW_i^V, \mathbf{r}_i) \end{aligned} \quad (4)$$

251 where $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_k}$ represents the parameter matrices,
252 $\mathbf{r}_i \in \mathbb{R}^{d_k \times d_w}$ is the relative position encodings of interaction features, d_w is the length of
253 features and i denotes the i -th attention head. d_{model} is the input feature dimension of the MHSA
254 layer. In this work, we use $d_{model} = 128$, $h = 4$, and $d_k = d_{model}/h = 32$.

255 As shown in the MHSA layer in **Figure 1c**, the input size is $W \times d$, which respectively represent
256 the length and encoding dimension of the MHC-II peptide interaction feature map. \mathbf{r}_i denote a
257 learnable relative position encodings matrix of the $head_i$, \mathbf{r} is the set of all \mathbf{r}_i . First, the input feature
258 X is multiplied by the weight matrices W_Q , W_K and W_V respectively to obtain matrices \mathbf{q} , \mathbf{k} and \mathbf{v} .
259 Then, matrix \mathbf{r} is multiplied by \mathbf{q} to get the *content-position* which forms part of the attention.
260 Following traditional self-attention mechanism, the product of \mathbf{q} and \mathbf{k} is calculated as *content-content*. By adding corresponding elements from the *content-position* and *content-content*, and then
261 applying the softmax function, we calculate the final attention weights. Finally, these attention
262

263 weights are multiplied by the matrix \mathbf{v} to obtain the output feature map of the self-attention layer with
264 position information.

265

266 **Supervised contrastive learning**

267 ***Data preprocessing for supervised contrastive pre-training.*** In the supervised contrastive learning
268 pre-training phase, we split the training dataset, BD2016, into ten distinct categories at intervals of
269 0.1 based on the transformed affinity score. **Figure 1e** provides a detailed illustration of this
270 categorization — for instance, category 0 includes affinity values within the range of [0.0, 0.1].
271 Various numbers of categories were trialed in our experiment; however, we found that neither too
272 many nor too few was beneficial. Particularly, having a high number of categories led to fewer
273 samples per group, which hindered the learning of discriminative features. Conversely, when there
274 were too few categories available, it would prevent the full realization of the substantial potential in
275 supervised contrastive learning. Therefore, we finally chose to retain the categorization number of
276 ten.

277 ***Supervised contrastive learning pre-training.*** Supervised contrastive learning combines the
278 strengths of both supervised and contrastive learning, achieving remarkable success across various
279 domains. It demonstrates powerful feature extraction abilities in computer vision tasks like image
280 classification [25] and semantic segmentation [26]. In natural language processing tasks such as text
281 classification [27] and sentiment analysis [28], this methodology excels in capturing nuances and
282 commonalities among texts, leading to improved model performance. In transfer learning [29]
283 scenarios, models enhanced by supervised contrastive learning exhibit robust generalization
284 capabilities, enabling efficient adaptation to new tasks. In this study, we innovatively apply supervised
285 contrastive learning to the regression prediction task of MHC-II peptide affinity prediction. By
286 leveraging the supervised contrastive loss, our approach can effectively identify the similarities and
287 dissimilarities in the MHC-II peptide interaction characteristics, resulting in significant
288 improvements in the predictive performance.

289 Next, we briefly introduce the main components of the supervised contrastive learning
290 framework and the expression equation of the supervised contrastive loss function.

291 We mainly divide the supervised contrastive learning framework into two parts: Encoder

292 Network and Projection Network. The specific architecture of the encoder network is described in
293 detail in ***backbone encoder network*** section. We represent the output features of the encoder network
294 as ***rep***. Following the encoding network, we employ a multi-layer projection network to map ***rep*** to
295 vector ***z***. The projection network is represented by the following formula:

296
$$\mathbf{z} = Proj(\mathbf{rep}) \in \mathbb{R}^{D_P} \quad (5)$$

297 Here, $Proj(\cdot)$ consists of two fully-connected layers and a ReLU activation function subsequently
298 normalizing the final output to the unit hypersphere in \mathbb{R}^{D_P} (where $D_P = 16$ in this study). In the end,
299 we utilize the inner product between ***z*** as a metric for their distance in the projected space, and
300 calibrate the model parameters via the contrastive loss, which is expressed as follows:

301
$$\mathcal{L} = \sum_{i \in I} \mathcal{L}_i = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp\left(z_i \cdot \frac{z_p}{\tau}\right)}{\sum_{a \in A(i)} \exp\left(z_i \cdot \frac{z_a}{\tau}\right)} \quad (6)$$

302 In a given batch, let $\{(\mathbf{x}_k, \mathbf{y}_k), label_k\}_{k=1,2,\dots,N}$ be peptide, MHC-II, label triplets and $\mathbf{z}_k =$
303 $Proj(Enc(\mathbf{x}_k, \mathbf{y}_k)) \in \mathbb{R}^{D_P}$. Let $i \in I \equiv \{1, 2, \dots, N\}$ be the index of a sample, typically referred to as
304 the ‘anchor’. $A(i) \equiv I \setminus \{i\}$ represents the set of indices not included in set I , $P(i) \equiv \{p \in$
305 $A(i) : label_p = label_i\}$ is the set of indices for all positives in the batch excluding i , and $|P(i)|$ is its
306 cardinality. The \cdot stands for the inner (dot) product, while $\tau \in \mathbb{R}^+$ is a scalar representing a
307 temperature parameter. A diagram of the supervised contrastive loss is shown in **Figure 1d**, where
308 the distance of MHC-peptide pairs within the anchor category is pulled closer.

309 Upon completing the supervised contrastive learning phase, we discard the projection network
310 $Proj(\cdot)$. In the subsequent fine-tuning stage, we employ the Mean Squared Error (MSE) loss function
311 to perform further model training based on the established encoder $Enc(\cdot)$ parameters.

312

313 **Two stages training process**

314 In the pre-training stage, we mapped the interaction feature ***rep***, with a dimension of 256, to a
315 representation vector ***z***, with a dimension of 16, via the projection network $Proj(\cdot)$. We optimized
316 the model parameters with the supervised contrastive loss and larger batch sizes (set to 512) for model
317 stability and speed. In the fine-tuning stage, we replaced the projection network with a linear layer
318 and sigmoid function, then optimized the pre-trained model with MSE loss. Additionally, we
319 determined the binding cores by identifying the highest-scoring positions from the output vector of

320 the bottleneck transformer during inference. A detailed description of the model implementation and
321 hyperparameter settings is provided in **Supplementary Text S2**.

322

323 RESULTS

324 Comparison of ConBoTNet and competing methods under 5-fold cross-validation

325 **Table 1** presents the average area under the ROC curve (AUC) and Pearson correlation coefficient
326 (PCC) of ConBoTNet and other competitive methods over all MHC-II molecules via 5-fold CV on
327 BD2016 data set. Since NetMHCIIpan-4.0 has already incorporated elution ligand mass spectrometry
328 data, it is not suitable to compare NetMHCIIpan-4.0 through two validation methods of 5-fold CV
329 and LOMO. Our proposed model, ConBoTNet, outperforms existing models in terms of both average
330 AUC and PCC. Specifically, ConBoTNet achieved an AUC of 0.865 and a PCC of 0.709, exceeding
331 existing SOTA methods. The AUC and PCC scores of the competing methods on 5-fold CV are as
332 follows: DeepMHCII (0.856, 0.690), NetMHCIIpan-3.2 (0.847, 0.679), PUFFIN (0.846, 0.676) and
333 DeepSeqPanII (0.741, 0.488). Moreover, the boxplots in **Figure 2a** clearly illustrate the overall
334 performance of each method across 61 MHC-II molecules, where ConBoTNet outperforms all other
335 methods. The asterisk (*) featured in the boxplot indicates the statistical significance of ConBoTNet's
336 performance compared to other methods, with four asterisks (****) indicating $P\text{-value} < 0.0001$.
337 **Figure 2b** provides a comparison between ConBoTNet and the current state-of-the-art method,
338 DeepMHCII, in terms of AUC and PCC values for each MHC-II molecule. Each dot in **Figure 2b**
339 corresponds to one MHC-II molecule. Specifically, the scatter plots reveal that ConBoTNet exhibited
340 varying degrees of performance improvement over 57 out of 61 MHC-II molecules (representing
341 ~93.4% of molecules) compared to DeepMHCII in terms of AUC and PCC.

342

343 Comparison of ConBoTNet and competing methods under LOMO

344 **Table 1** also shows that ConBoTNet transcended other competing methods in terms of average AUC
345 and PCC under LOMO on BD2016 data set. The experimental results are consistent with 5-fold CV.
346 Specifically, ConBoTNet achieved the highest AUC and PCC scores of 0.789 and 0.568, respectively.
347 Furthermore, the boxplots in **Figure 2c** demonstrate that ConBoTNet outperforms competing

348 methods in most cases. ConBoTNet shows statistically significant improvement over the suboptimal
349 method, DeepMHCII (P-value = 0.0413). Additionally, **Figure 2d** illustrates that ConBoTNet
350 outperforms the suboptimal method for approximately 72.1% and 73.8% of the MHC-II molecules
351 based on the AUC and PCC metrics, respectively.

352

353 **Comparison of ConBoTNet and competing methods on independent testing set**

354 The BD2023 dataset was collected from the latest IEDB release by the MHC-II automated benchmark
355 platform. Any overlapping items with BD2016 within BD2023 were subsequently removed. We then
356 divided this dataset into two subsets: BD2023 (binary) and BD2023 (IC_{50}), based on the types of
357 measurement data. **Table 1** shows the evaluation results of each model on BD2023. For
358 NetMHCIIpan-3.2 and NetMHCIIpan-4.0, these results were obtained from the benchmark testing
359 platform. Meanwhile, the results for DeepSeqPanII, DeepMHCII, and ConBoTNet were obtained by
360 training these models on BD2016 and then tested on this independent dataset. Unfortunately, we could
361 not provide test results on BD2023 for the PUFFIN method, as the training code was not disclosed,
362 and the author-provided model download link was invalid. According to **Table 1**, our model has
363 shown improvements in terms of average AUC compared to the current best-performance models.
364 Specifically, there was a 2.7% improvement (AUC, 0.637 versus 0.620) in BD2023 (binary) and a
365 1.7% improvement (AUC, 0.845 versus 0.831) in BD2023 (IC_{50}). The distributions of AUC scores
366 for each MHC-II molecules on both subsets are further depicted in **Figure 2e**, where ConBoTNet
367 demonstrates statistically significant improvements compared to other models (paired t-test, P-value
368 = 0.0095 vs. NetMHCIIpan-4.0 and P-value = 0.0467 vs. DeepMHCII). Additionally, **Figure 2f**
369 shows that ConBoTNet outperforms the suboptimal methods (NetMHCIIpan-4.0 and DeepMHCII)
370 for approximately 91.0% and 76.5% of the MHC-II molecules in the two BD2023 subsets,
371 respectively.

372

373 **Comparison of ConBoTNet and single model based on ensemble methods**

374 Various algorithms can enhance binding affinity prediction by ensembling multiple models of the
375 same or similar architecture. For example, NetMHCIIpan-3.2 uses 40 neural networks for ensemble;
376 PUFFIN is a deep ensemble model composed of 20 deep residual convolutional neural networks;

377 DeepMHCII is an ensemble of 20 models using a simple average method. In this section, to examine
378 the possibility of performance improvement by relying solely on the model architecture itself, we
379 discuss the comparison between ConBoTNet and a single model of existing ensemble algorithms.
380 However, we did not include PUFFIN in the performance comparison since it did not provide neither
381 implementation code nor the trained model. We compared ConBoTNet with the SOTA affinity
382 prediction method, DeepMHCII, which outperformed PUFFIN (discussed in detail in DeepMHCII
383 by You *et al.*), and DeepSeqPanII, a single-model algorithm. The performance comparison results on
384 5-fold CV test for 61 MHC-II molecules are shown in **Figure 3**, while the overall performance
385 evaluation metrics are presented in **Table S2**. As can be seen, ConBoTNet outperformed DeepMHCII
386 on about 95.1% of MHC-II molecules with statistical significance (P-value < 0.0001). Specifically,
387 ConBoTNet achieved the best average AUC of 0.837 on 5-fold CV, which was 1.9% higher than the
388 SOTA method, DeepMHCII (0.821), and 13.0% higher than DeepSeqPanII (0.741), respectively.
389 Similarly, the average PCC was also improved compared to DeepMHCII (increased by 4.6%, 0.657
390 versus 0.628), indicating the excellent performance of the ConBoTNet model. Additionally, a detailed
391 discussion of the ablation experiments about bottleneck transformer modules and supervised
392 contrastive learning is provided in **Supplementary Text S3**.

393

394 **Binding core prediction & antigen weight visualization**

395 The final column in **Table 1** shows the ability of ConBoTNet and the compared methods to predict
396 the binding core on BC2015. It suggests that both ConBoTNet and NetMHCIIpan-3.2 achieved
397 remarkable prediction accuracy, correctly identifying 45 out of 51 MHC-II peptide complexes. **Figure**
398 **4a** provides a detailed visualization of the results generated by ConBoTNet. Accurate predictions are
399 denoted in blue, while incorrect predictions are highlighted in red. To better understand the model's
400 learning capability and potential error sources, we utilized the Captum [30] tool to analyze the
401 attention weights assigned by the ConBoTNet model to different positions of the peptide sequences.
402 These results are visualized through a heatmap in **Figure 4a**. As the binding core consists of a
403 continuous fragment of nine amino acids within the peptide sequence, we demarcate this fragment
404 using a box encompassing nine amino acids with the highest sum of interest. The accurately and
405 inaccurately predicted binding cores are marked by blue boxes and red boxes, respectively. Excitingly,

406 ConBoTNet demonstrates a strong focus on the most critical aspect of peptide binding affinity, i.e.,
407 the binding core. For some MHC-peptide complexes, such as PDB IDs 1UVQ and 1IAO, the
408 ConBoTNet model may have failed to identify the real binding cores due to the influence of the
409 peptides' flanking regions [23]. Nevertheless, this observation underscores the model's capability to
410 emphasize the peptide fragment that is most important for the affinity effect. In **Figure 4b**, we present
411 five PDB structures depicting the peptide-binding grooves of human MHC-II (HLA-DR, DP, DQ)
412 and mouse MHC-II (H-2) molecules. Amino acids with the highest attention weights are usually
413 located at the MHC peptide-binding pockets and involved in the polar interactions between the
414 peptide and MHC-II molecule or TCR, e.g. In the HLA-DR2b/MBP-peptide complex (PDB ID:
415 1YMM), a bulky amino acid F8 of the peptide forms hydrogen bonds with Q9 and N62 at the MHC-
416 II α chain. This also illustrates that our proposed ConBoTNet can identify binding anchors that
417 interact with MHC-II or TCR.

418

419 **Sequence logo representations**

420 We visualized the binding motifs of MHC-II molecules, identified by ConBoTNet and DeepMHCII,
421 as sequence logos using Seq2Logo v2.0 [31]. Following the description in You *et al.* [17], we initially
422 calculated the binding scores for 100,000 random peptides from UniProt [32] in relation to a specific
423 MHC-II molecule. Subsequently, we selected the top 1% of peptides with the highest binding scores
424 to generate the sequence logos under the default settings.

425 Furthermore, we focused on six MHC-II molecules that demonstrated high predictive accuracy
426 on the BD2023 test set and had detailed motif information in the SYFPEITHI database [33]. The
427 binding groove of the MHC-II molecule is open at both ends, enabling the interaction with peptides
428 of variable lengths. The ends of the binding pocket are called the major anchors (P1, P9) from the N-
429 terminus to the C-terminus, whereas the smaller pockets generate auxiliary anchors (P4, P6, P7) [34].
430 **Figure 5** illustrates the sequence logos predicted by two advanced methods, namely ConBoTNet and
431 DeepMHCII, for six MHC-II molecules with diverse anchor types including P1-P4-P6-P9, P1-P4-P7
432 and P1-P4-P9.

433 Next, we assessed the anchor recognition capabilities of ConBoTNet and DeepMHCII. For the
434 molecules DRB1*01:01, DRB1*04:05, DRB1*07:01, and DRB1*13:02, both methods could identify

435 their anchors, namely P1-P4-P6-P9. However, the distinction between pockets became less
436 pronounced due to the integration of contextual information by the multi-head attention module.
437 Consequently, the anchors identified by ConBoTNet were not as distinct as those provided by
438 DeepMHCII. In the case of DRB1*15:01 with P1-P4-P7 anchors, only ConBoTNet accurately
439 identified P7. Moreover, for DRB5*01:01 with P1-P4-P9 anchors, DeepMHCII failed to identify the
440 correct P4 anchor. Furthermore, we also analyzed the anchor amino acid preferences, whose details
441 are described in **Supplementary Text S4**. Overall, ConBoTNet achieved a more accurate predictive
442 performance for predicting the anchors and amino acid preferences than the competing methods.
443

444 **Supervised contrastive learning enhances affinity prediction**

445 In this section, we explore the role of supervised contrastive learning (pre-training) in improving
446 affinity prediction models. For more intuitive representations, we output the normalized embeddings
447 through the pre-trained model under 5-fold CV. Next, we use UMAP tool [35], a general-purpose
448 manifold learning and dimensionality reduction algorithm, to visualize these embeddings. **Figure 6**
449 and **Figure S1** show the visualization of the original distribution of the testing data (one-fold) and the
450 embedding visualization of the corresponding testing set under 5-fold CV. Leveraging the labeled
451 data, the supervised contrastive loss encourages the network to pull normalized embeddings from the
452 same class closer, while pushing embeddings from different classes apart. In the pre-training data
453 processing, we hard-divided the data into ten classes according to the affinity value, following **Figure**
454 **1e**. Interestingly, the model via supervised contrastive learning (pre-trained stage) can learn a gradient
455 “curve” that varies by affinity value. The robust base model established a solid foundation for the
456 subsequent fine-tuning phase. In addition, the base model can be used to fine-tune MHC-II molecular
457 data with fewer data, enhancing the affinity prediction accuracy for molecules with smaller samples.
458 It also demonstrates why fine-tuning the pre-trained model for only 1-2 epochs can produce excellent
459 prediction accuracy.
460

461 **CONCLUSIONS**

462 In this study, we have designed and developed ConBoTNet, a novel deep-learning algorithm for
463 predicting MHC-II peptide binding affinity. The bottleneck transformer module successfully captured

464 the binding core and precise anchor information of peptide MHC-II from multi-source interaction
465 features (including embedding and one-hot encoding), and then learns a robust feature representation
466 based on the similarities and differences between the interaction features through supervised
467 contrastive learning for high-precision affinity prediction.

468 Extensive experiments have been conducted on benchmark datasets, including 5-fold CV, LOMO,
469 independent test, and binding core prediction. As a result, ConBoTNet significantly outperformed all
470 five SOTA competing methods. Furthermore, ConBoTNet identified the binding cores as well as
471 important anchors more accurately than existing methods. In terms of model interpretability, we used
472 the Captum algorithm to explain the influence of binding core and peptides flanking regions on
473 affinity prediction by analyzing the attention weights on peptide sequences; and visually
474 demonstrated how supervised contrastive learning could enhance affinity prediction. All these results
475 highlight the utility and significance of ConBoTNet for relevant biomedical discovery and
476 interpretability. One limitation of our study is that we focused on the peptide-MHC binding prediction
477 rather than T cell epitope identification problem. However, peptide-MHC binding is a prerequisite for
478 T cell immunogenicity, for which multiple studies have shown that MHC peptide binding strength is
479 correlated strongly with peptide immunogenicity [36]. This study provides a useful groundwork for
480 our better understanding of the interaction conformation between T-cell receptors and epitopes. We
481 plan to integrate multivariate prior biological and biomedical knowledge and design highly
482 interpretable model architectures to develop advanced deep learning methods for predicting of T cell
483 receptor-epitope binding specificity [37, 38] in our future work.

484

485 FUNDING

486 This work was supported by the National Natural Science Foundation of China (62372234,
487 62072243), the Postgraduate Research & Practice Innovation Program of Jiangsu Province
488 (KYCX23_0490), the Australia Research Council (LP220200614) and the Major Inter-Disciplinary
489 Research (IDR) project awarded by Monash University.

490 DATA AVAILABILITY

491 Data that are involved in this work can be downloaded from GitHub at
492 <https://github.com/shenlongchen/conbotnet>.
493

494 **Key Points**

- 495 • We present a novel deep-learning method, termed ConBoTNet, for predicting the MHC-II
496 peptide binding affinity, binding core, and anchor positions within the binding core.
- 497 • ConBoTNet accurately identifies the binding core and anchors from multi-source interaction
498 features via supervised contrastive pre-training and bottleneck transformer modules.
- 499 • Extensive experiments on benchmark datasets, encompassing 5-fold CV, LOMO, independent
500 testing, and binding core prediction, reveal that ConBoTNet significantly outperforms all five state-
501 of-the-art methods.
- 502 • We utilize the Captum tool to scrutinize the model's assigned weights to peptide sequences,
503 thereby interpreting how binding cores impact binding affinity. Additionally, we employ UMAP to
504 visualize pre-trained features, demonstrating how supervised contrastive learning extracts
505 discriminative features for precise affinity prediction.

506

507 **Competing interests**

508 The authors declare that they have no competing interests.

509

510

511 REFERENCES

- 512 1. Alspach E, Lussier DM, Miceli AP et al. MHC-II neoantigens shape tumour immunity and response to
513 immunotherapy, *Nature* 2019;574:696-701.
- 514 2. Liu H, Wilson KR, Firth AM et al. Ubiquitin-like protein 3 (UBL3) is required for MARCH ubiquitination of
515 major histocompatibility complex class II and CD86, *Nature Communications* 2022;13:1934.
- 516 3. Sagan SA, Moinfar Z, Moseley CE et al. T cell deletional tolerance restricts AQP4 but not MOG CNS
517 autoimmunity, *Proceedings of the National Academy of Sciences* 2023;120:e2306572120.
- 518 4. Blass E, Ott PA. Advances in the development of personalized neoantigen-based therapeutic cancer
519 vaccines, *Nature Reviews Clinical Oncology* 2021;18:215-229.
- 520 5. Wang Y, Xiang Y, Xin VW et al. Dendritic cell biology and its role in tumor immunotherapy, *Journal of
521 hematology & oncology* 2020;13:1-18.
- 522 6. Ishina IA, Zakharova MY, Kurbatskaia IN et al. MHC Class II Presentation in Autoimmunity, *Cells*
523 2023;12:314.
- 524 7. Moore TV, Nishimura MI. Improved MHC II epitope prediction—a step towards personalized medicine,
525 *Nature Reviews Clinical Oncology* 2020;17:71-72.
- 526 8. Meraviglia-Crivelli D, Zheleva A, Barainka M et al. Therapeutic Strategies to Enhance Tumor Antigenicity:
527 Making the Tumor Detectable by the Immune System, *Biomedicines* 2022;10:1842.
- 528 9. Kisielow J, Obermair F-J, Kopf M. Deciphering CD4+ T cell specificity using novel MHC–TCR chimeric
529 receptors, *Nature immunology* 2019;20:652-662.
- 530 10. Chang ST, Ghosh D, Kirschner DE et al. Peptide length-based prediction of peptide–MHC class II binding,
531 *Bioinformatics* 2006;22:2761-2767.
- 532 11. Jones EY, Fugger L, Strominger JL et al. MHC class II proteins and disease: a structural perspective, *Nature
533 Reviews Immunology* 2006;6:271-282.
- 534 12. Finotello F, Rieder D, Hackl H et al. Next-generation computational tools for interrogating cancer immunity,
535 *Nature Reviews Genetics* 2019;20:724-746.
- 536 13. Soleymani S, Tavassoli A, Housaindokht MR. An overview of progress from empirical to rational design in
537 modern vaccine development, with an emphasis on computational tools and immunoinformatics approaches,
538 *Computers in biology and medicine* 2022;140:105057.
- 539 14. Fatima I, Ahmad S, Abbasi SW et al. Designing of a multi-epitopes-based peptide vaccine against rift valley
540 fever virus and its validation through integrated computational approaches, *Computers in biology and medicine*
541 2022;141:105151.
- 542 15. Jensen KK, Andreatta M, Marcatili P et al. Improved methods for predicting peptide binding affinity to MHC
543 class II molecules, *Immunology* 2018;154:394-406.
- 544 16. Zeng H, Gifford DK. Quantification of uncertainty in peptide-MHC binding prediction improves high-
545 affinity peptide selection for therapeutic design, *Cell systems* 2019;9:159-166. e153.
- 546 17. You R, Qu W, Mamitsuka H et al. DeepMHCII: a novel binding core-aware deep interaction model for
547 accurate MHC-II peptide binding affinity prediction, *Bioinformatics* 2022;38:i220-i228.
- 548 18. Venkatesh G, Grover A, Srinivasaraghavan G et al. MHCAttnNet: predicting MHC-peptide bindings for MHC
549 alleles classes I and II using an attention-based deep neural model, *Bioinformatics* 2020;36:i399-i406.
- 550 19. Liu Z, Jin J, Cui Y et al. DeepSeqPanII: an interpretable recurrent neural network model with attention
551 mechanism for peptide-HLA class II binding prediction, *IEEE/ACM Transactions on Computational Biology and
552 Bioinformatics* 2021;19:2188-2196.
- 553 20. Vita R, Mahajan S, Overton JA et al. The immune epitope database (IEDB): 2018 update, *Nucleic acids*

- 554 research 2019;47:D339-D343.
- 555 21. Andreatta M, Trolle T, Yan Z et al. An automated benchmarking platform for MHC class II binding
- 556 prediction methods, Bioinformatics 2018;34:1522-1528.
- 557 22. Karosiene E, Rasmussen M, Blicher T et al. NetMHCIIpan-3.0, a common pan-specific MHC class II
- 558 prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ,
- 559 Immunogenetics 2013;65:711-724.
- 560 23. Holland CJ, Cole DK, Godkin A. Re-directing CD4+ T cell responses with the flanking residues of MHC class
- 561 II-bound peptides: the core is not enough, Frontiers in immunology 2013;4:172.
- 562 24. Srinivas A, Lin T-Y, Parmar N et al. Bottleneck transformers for visual recognition. In: Proceedings of the
- 563 IEEE/CVF conference on computer vision and pattern recognition. 2021, p. 16519-16529.
- 564 25. Wang P, Han K, Wei X-S et al. Contrastive learning based hybrid networks for long-tailed image
- 565 classification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021, p.
- 566 943-952.
- 567 26. Xie J, Xiang J, Chen J et al. C2am: Contrastive learning of class-agnostic activation map for weakly
- 568 supervised object localization and semantic segmentation. In: Proceedings of the IEEE/CVF Conference on
- 569 Computer Vision and Pattern Recognition. 2022, p. 989-998.
- 570 27. Chen J, Zhang R, Mao Y et al. Contrastnet: A contrastive learning framework for few-shot text classification.
- 571 In: Proceedings of the AAAI Conference on Artificial Intelligence. 2022, p. 10492-10500.
- 572 28. Hu X, Lin L, Liu A et al. A multi-level supervised contrastive learning framework for low-resource natural
- 573 language inference, IEEE/ACM Transactions on Audio, Speech, and Language Processing 2023.
- 574 29. Sun J, Wei D, Ma K et al. Boost Supervised Pretraining for Visual Transfer Learning: Implications of Self-
- 575 Supervised Contrastive Representation Learning. In: Proceedings of the AAAI Conference on Artificial
- 576 Intelligence. 2022, p. 2307-2315.
- 577 30. Kokhlikyan N, Miglani V, Martin M et al. Captum: A unified and generic model interpretability library for
- 578 pytorch, arXiv preprint arXiv:2009.07896 2020.
- 579 31. Thomsen MCF, Nielsen M. Seq2Logo: a method for construction and visualization of amino acid binding
- 580 motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of
- 581 amino acid enrichment and depletion, Nucleic acids research 2012;40:W281-W287.
- 582 32. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK et al. SignalP 5.0 improves signal peptide predictions
- 583 using deep neural networks, Nature biotechnology 2019;37:420-423.
- 584 33. Schuler MM, Nastke M-D, Stevanović S. SYFPEITHI: database for searching and T-cell epitope prediction,
- 585 Immunoinformatics: Predicting immunogenicity in silico 2007;75-93.
- 586 34. Ferrante A. HLA-DM: arbiter conformationis, Immunology 2013;138:85-92.
- 587 35. Becht E, McInnes L, Healy J et al. Dimensionality reduction for visualizing single-cell data using UMAP,
- 588 Nature biotechnology 2019;37:38-44.
- 589 36. Kim JY, Bang H, Noh S-J et al. DeepNeo: a webserver for predicting immunogenic neoantigens, Nucleic
- 590 acids research 2023:gkad275.
- 591 37. Lu T, Zhang Z, Zhu J et al. Deep learning-based prediction of the T cell receptor–antigen binding specificity,
- 592 Nature machine intelligence 2021;3:864-875.
- 593 38. Peng X, Lei Y, Feng P et al. Characterizing the interaction conformation between T-cell receptors and
- 594 epitopes with deep learning, Nature machine intelligence 2023;5:395-407.

595

Figure Legends

Figure 1. Model architecture of ConBoTNet. a, Schematic diagram of the binding of MHC-II molecules and peptides on the surface of the antigen-presenting cell. b, The architecture of ConBoTNet and the training pipeline. c, Detailed structure of the multi-head self-attention (MHSA) layer. d, Schematic diagram of supervised contrastive loss. e, Based on the division of binding affinity values of BD2016, the data processing flow of constructing the pre-training classification dataset.

Figure 2. Performance evaluation of ConBoTNet and other competing methods on different benchmarks. a, Boxplots of the performance of ConBoTNet method and competing methods under 5-fold CV. b, The two scatter plots of point-to-point comparisons between ConBoTNet and the current best competing method on 5-fold CV under the AUC and PCC metrics, respectively. c, Boxplots of the ConBoTNet method and competing methods on leave-one-molecule-out (LOMO). d, The two scatter plots of point-to-point comparisons between ConBoTNet and the current best competing method on LOMO under the AUC and PCC metrics, respectively. e, Boxplots of the performance of ConBoTNet method and competing methods on BD2023 data set. f, The two scatter plots of point-to-point comparisons between ConBoTNet and the current best competing method on the BD2023 subset binary and IC_{50} under the AUC index, respectively.

Figure 3. Performance evaluation of ConBoTNet and competing methods on a single model under 5-fold CV. a. Boxplot representing the performance of ConBoTNet, DeepSeqPanII, and DeepMHCII on 5-fold CV under the AUC and PCC metrics. b, c, The scatter plots of point-to-point comparisons between ConBoTNet and the current best competing method on 5-fold CV under the AUC and PCC metrics, respectively.

Figure 4. Visualization of ConBoTNet average attention to antigen encodings on the BC2015. a. Heatmap visualization of each position of the antigenic peptide sequences of 51 MHC-II peptide complexes on BD2015. Each row in the heatmap represents a visualization of the average attention for each position of the antigen sequence of the complex, and each column denotes the relative position. Each row of the “true cores” column is the corresponding peptide sequence and the binding core is marked in blue. Meanwhile, each row of the “predicted cores” column represents the binding core position predicted by the ConBoTNet method. A correct prediction is colored in blue, whereas an incorrect one is marked red. b. Mark the high attention peptide binding sites (green) on the peptide MHC-II complex structure, along with the polar interactions (blue) with MHC-II (gray) and TCR (pink).

Figure 5. Sequence logo representations by ConBoTNet and DeepMHCII. Each sequence logo has nine positions (pockets) on the horizontal axis. At each position, the total height represents the relative information content of the corresponding position within the motif. Additionally, the height of each letter indicates the frequency of the corresponding amino acid in the position.

Figure 6. Visualization of the original distribution of the testing data (one-fold) and the embedding visualization of the corresponding testing set under 5-fold CV. a, b, and c respectively correspond

to the dimension reduction visualization of the original data of the 1st-fold, 2nd-fold, and 3rd-fold. D, e, and f respectively represent the embedding visualizations learned by the corresponding fold in the pre-training model.

Table Legends

Table 1. Performance comparison of ConBoTNet and other competing methods.

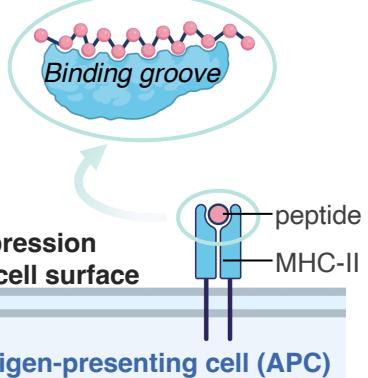
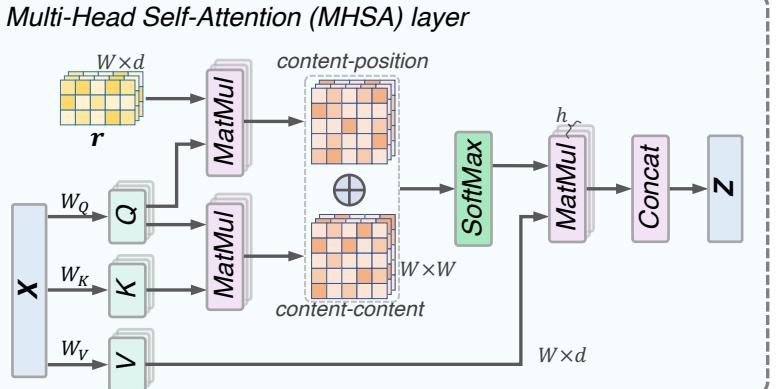
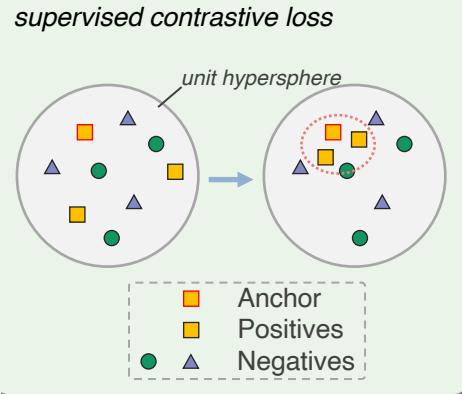
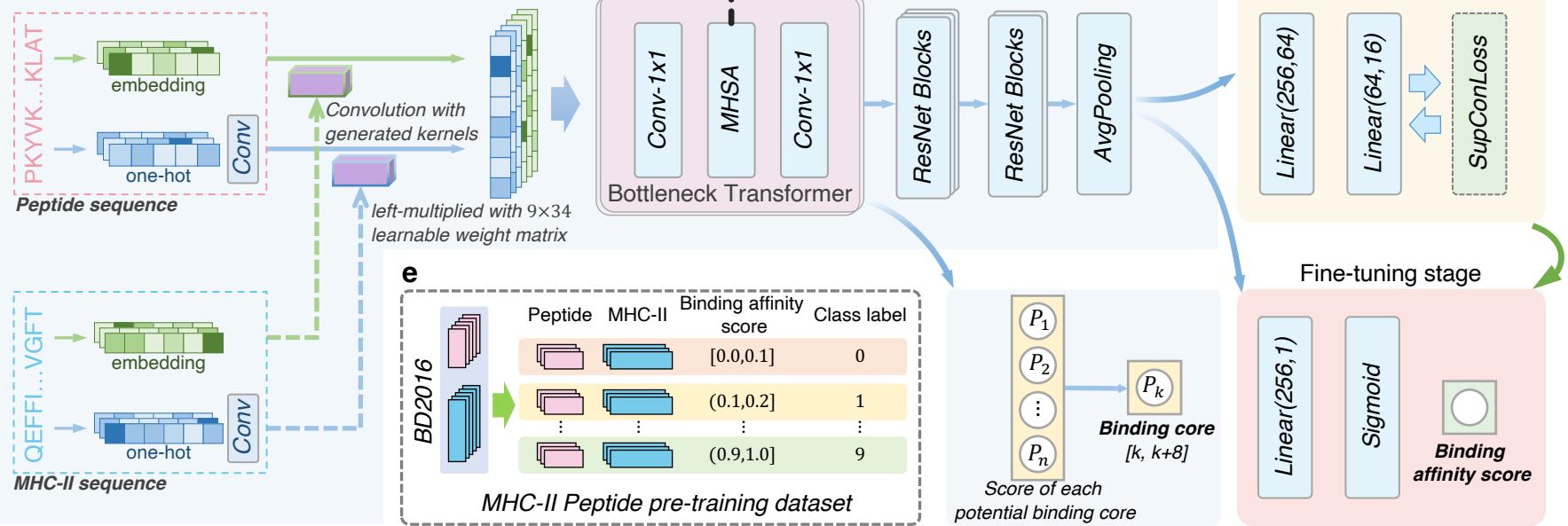
Supplementary Table legends

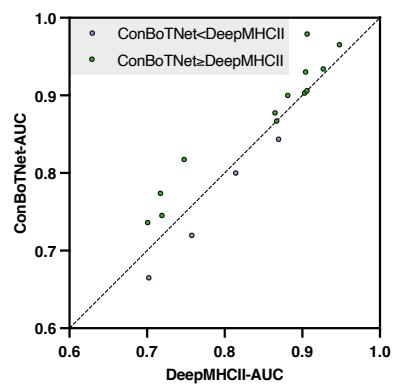
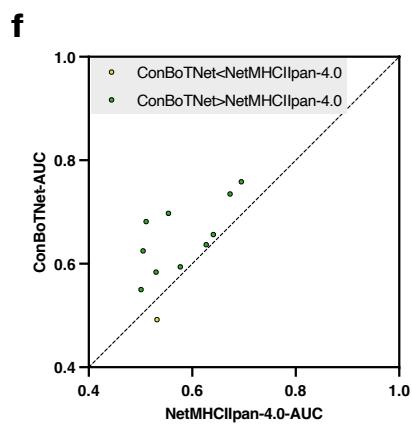
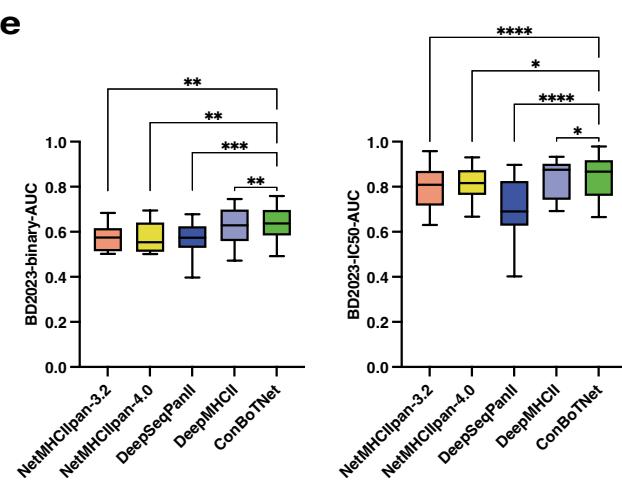
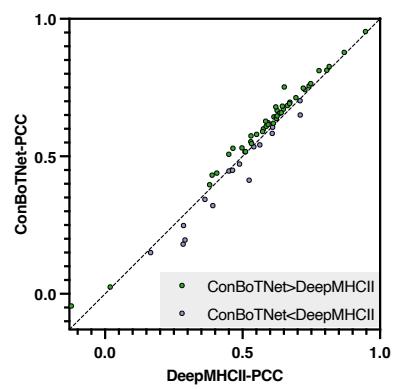
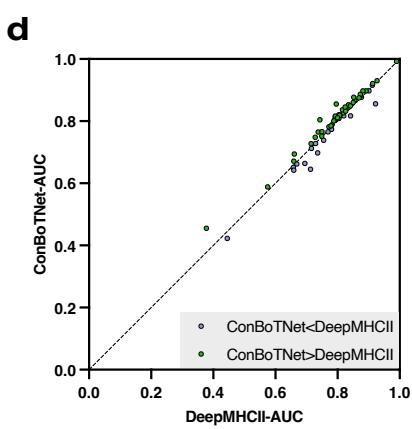
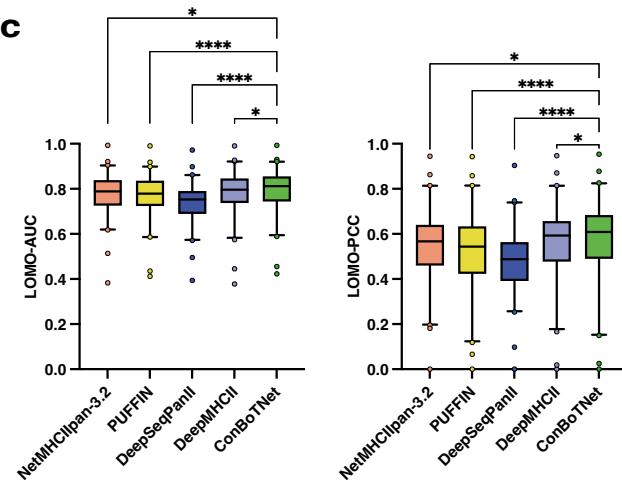
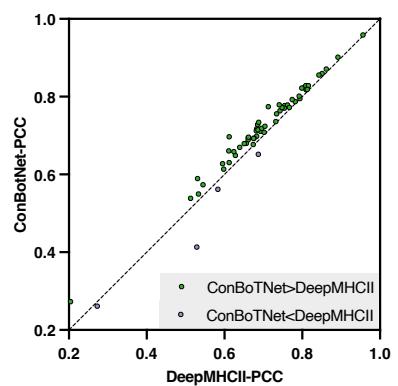
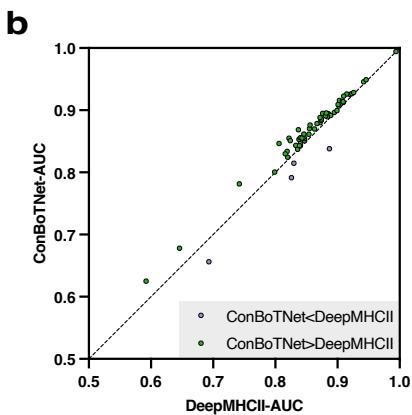
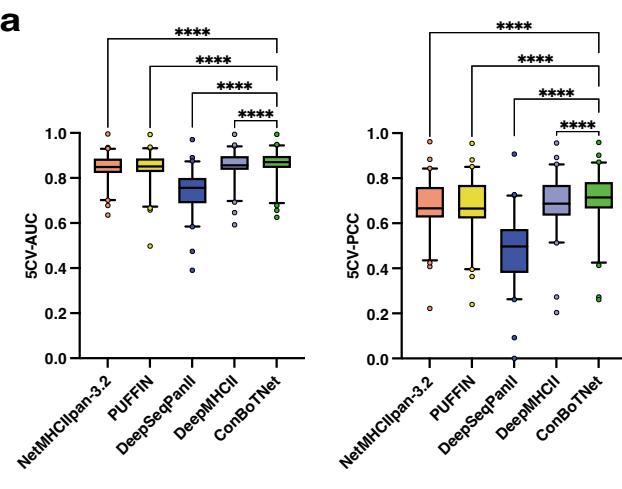
Table S1. Hyperparameters of ConBoTNet and the corresponding search space

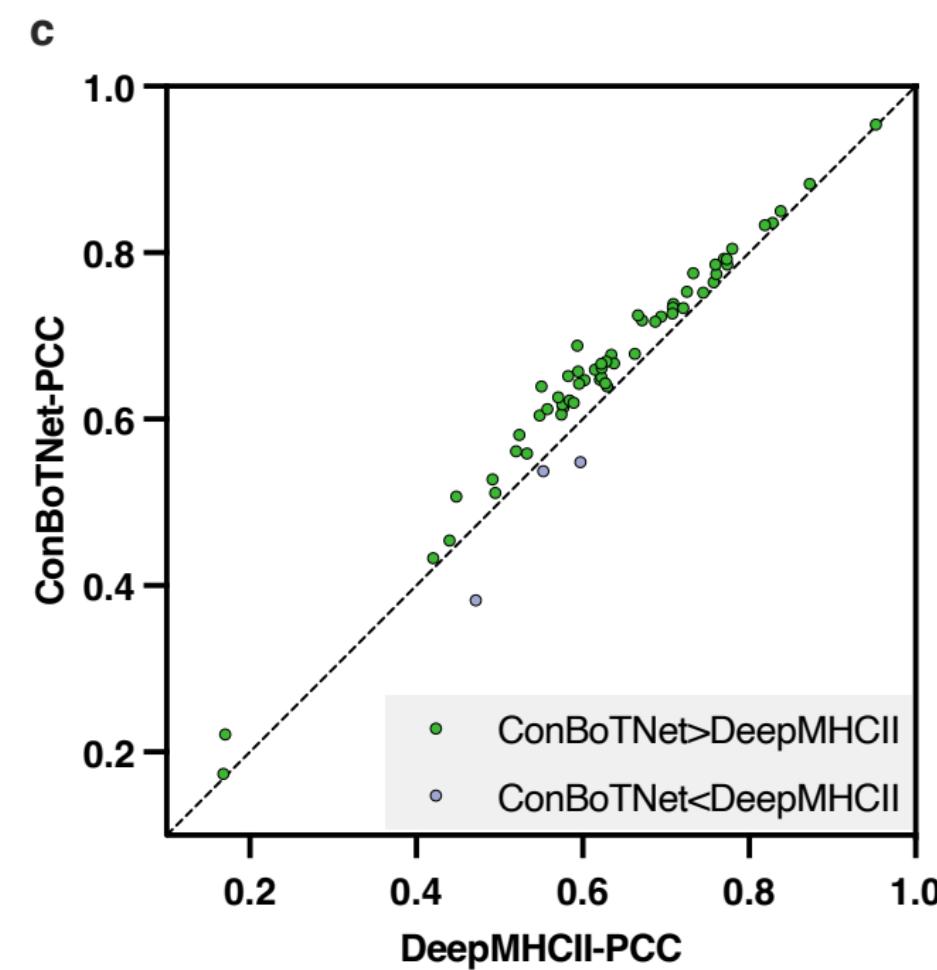
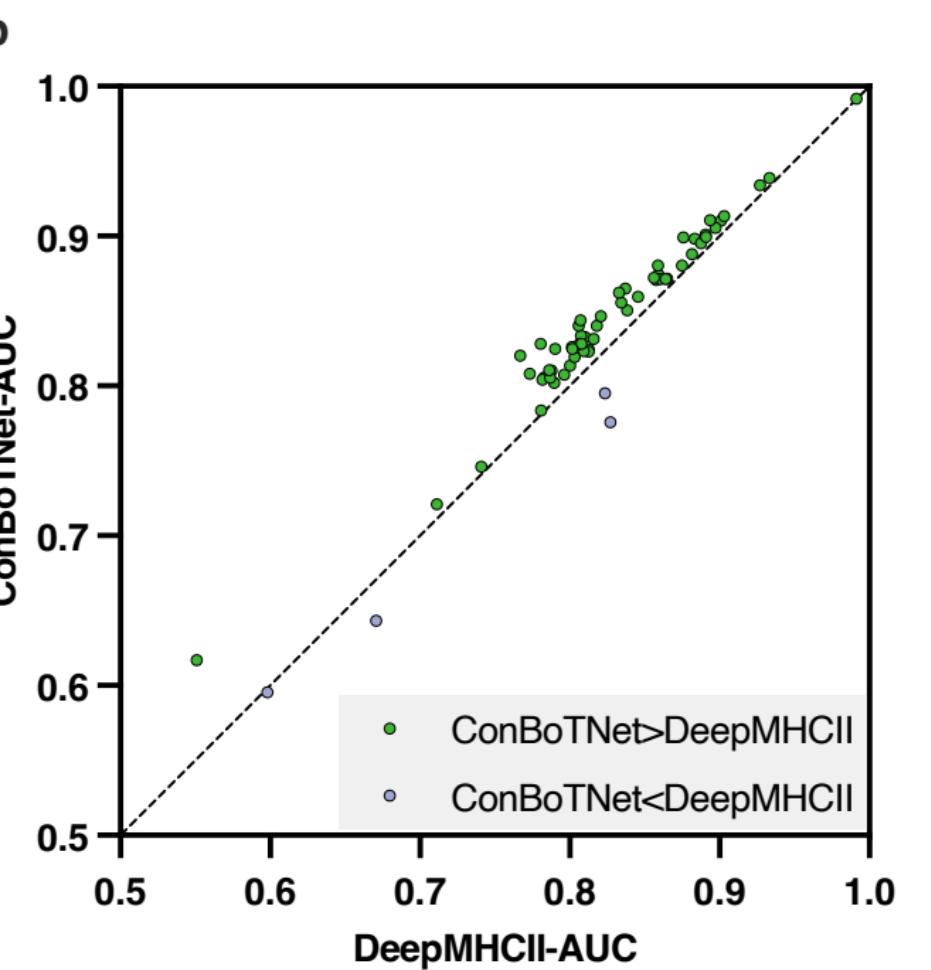
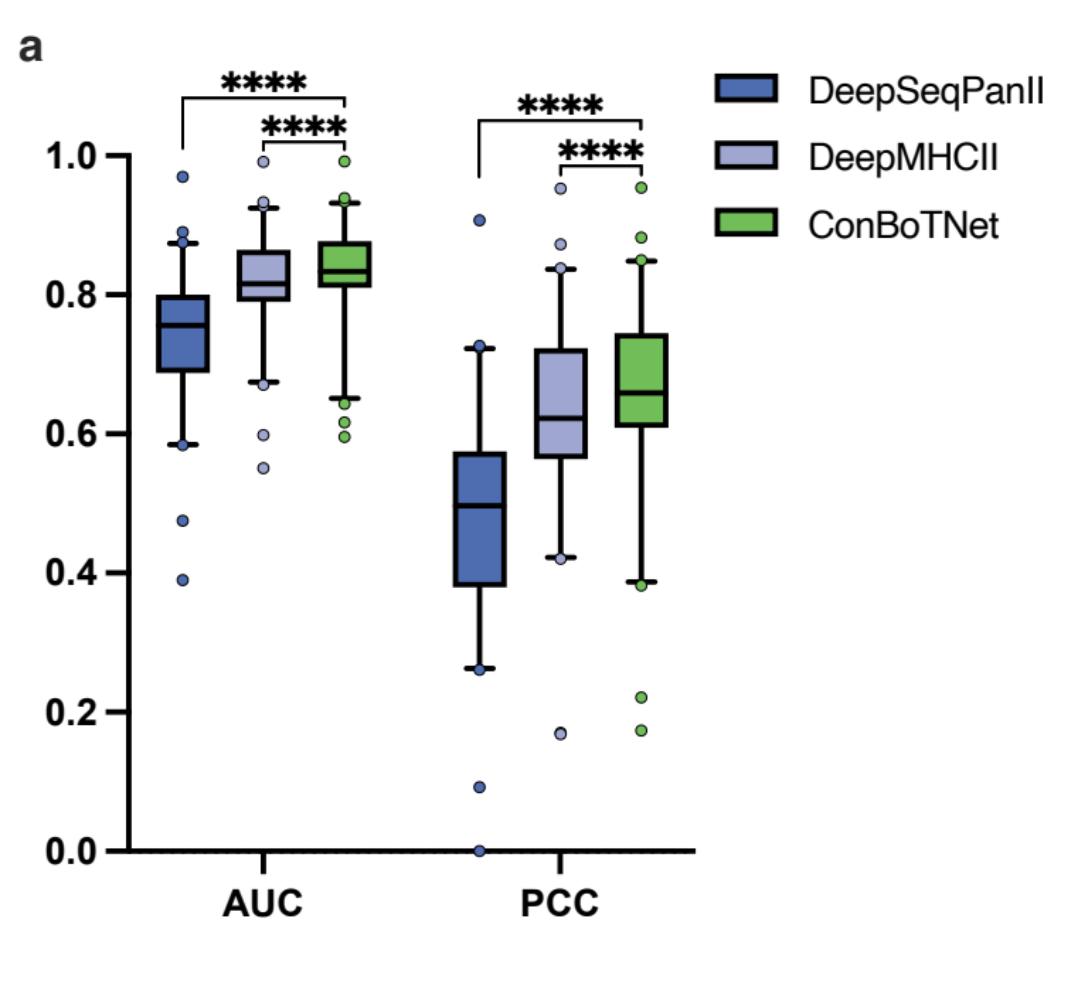
Table S2. Performance of ConBoTNet and competing methods on single model

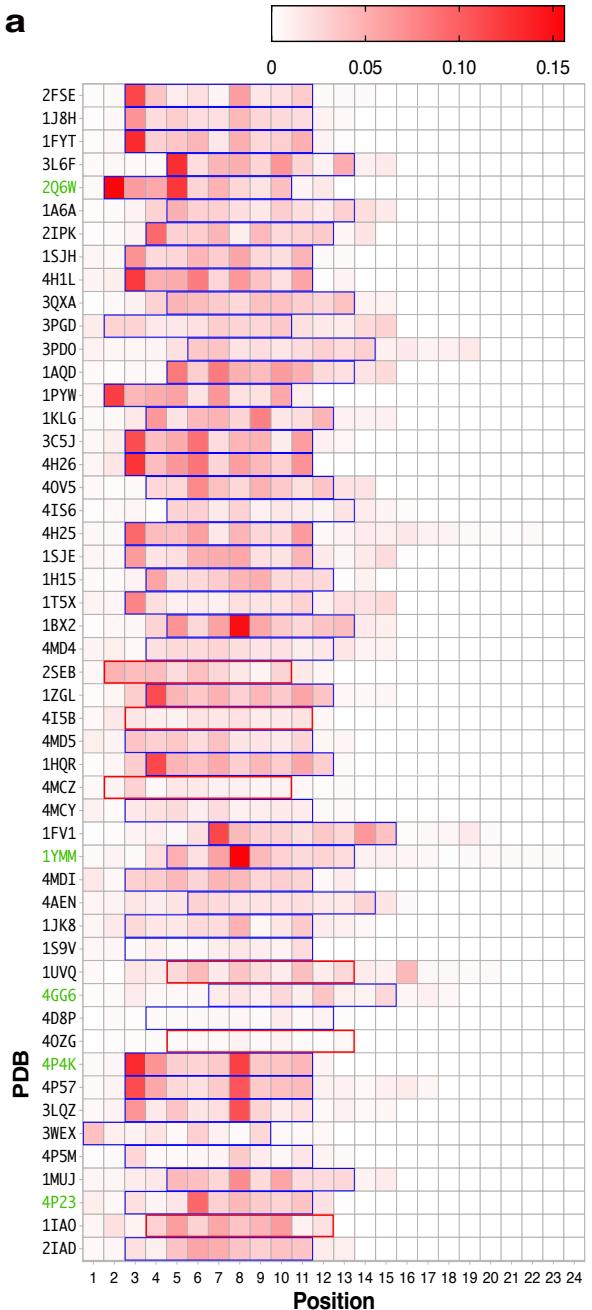
Supplementary Figure Legends

Figure S1. Visualization of the original distribution of the testing data (one-fold) and the embedding visualization of the corresponding testing set under 5-fold CV. a and b respectively correspond to the dimension reduction visualization of the original data of the 4th-fold and 5th-fold. c and d respectively represent the embedding visualizations learned by the corresponding fold in the pre-training model.

a**c****d****b**



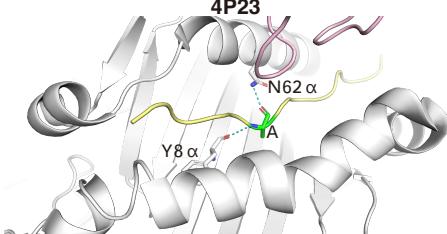
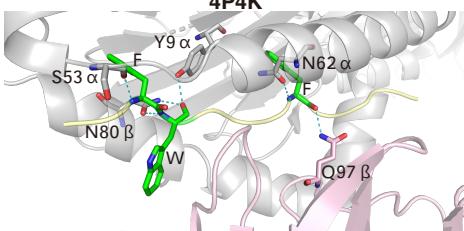
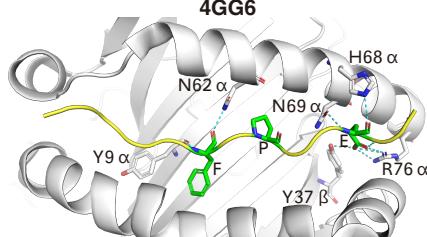
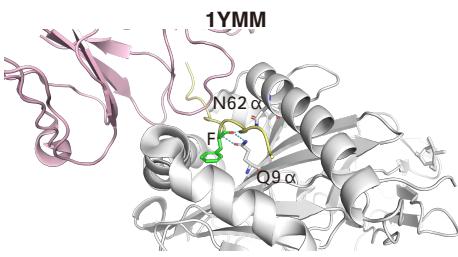
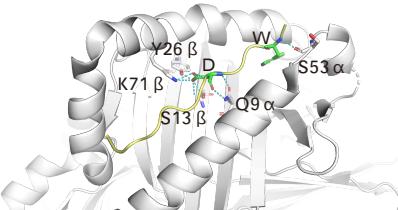




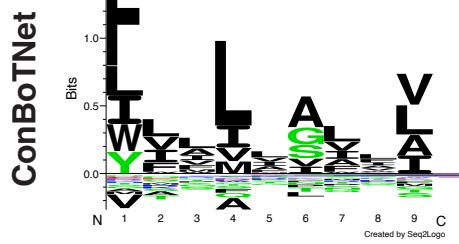
True cores

Predicted cores

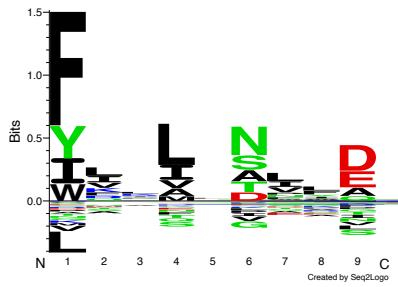
b



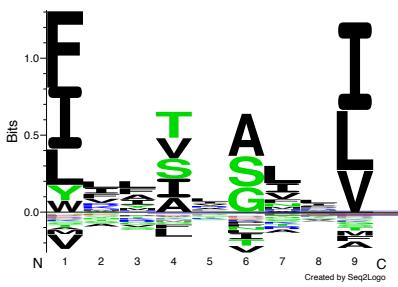
DRB1*01:01



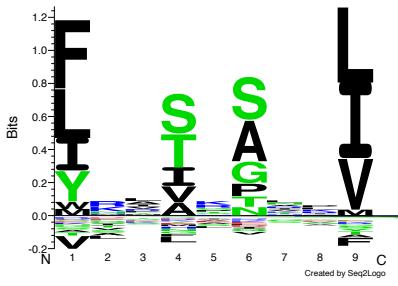
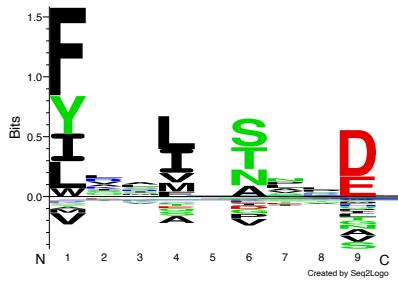
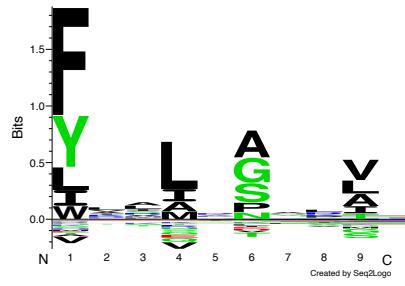
DRB1*04:05



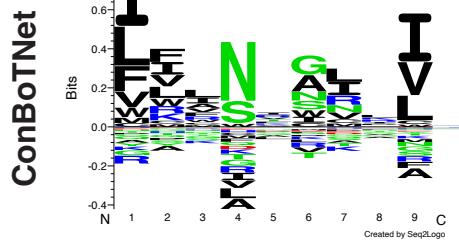
DRB1*07:01



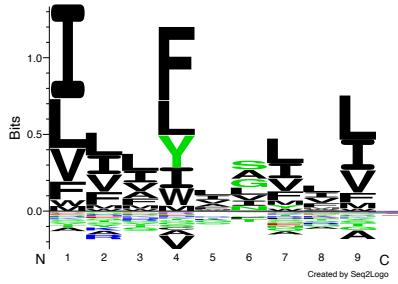
DeepMHCI



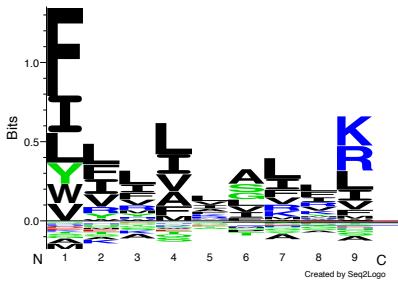
DRB1*13:02



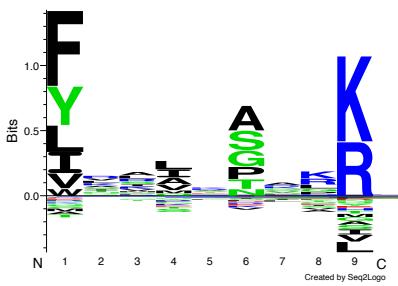
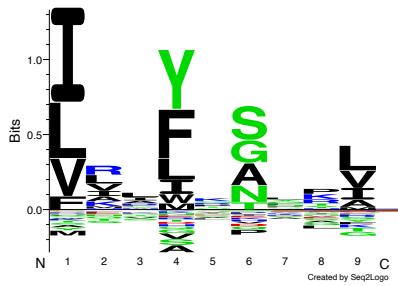
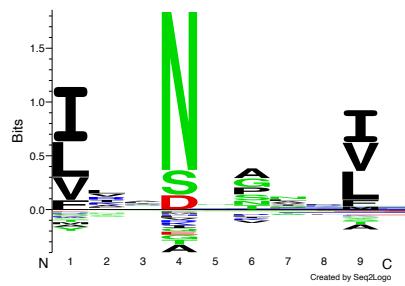
DRB1*15:01



DRB5*01:01



DeepMHCI



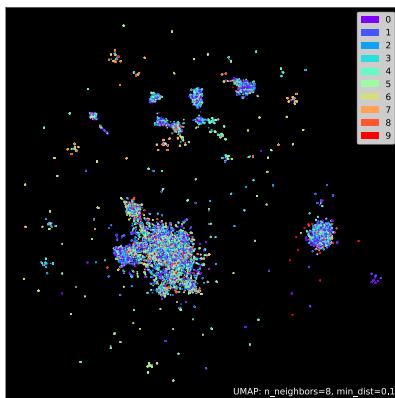
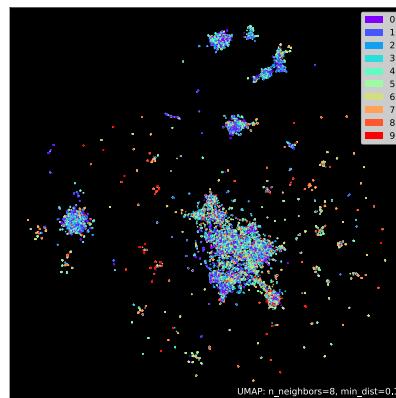
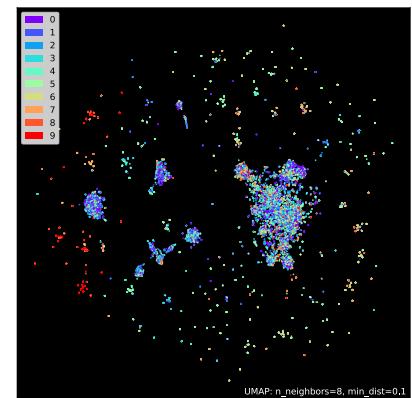
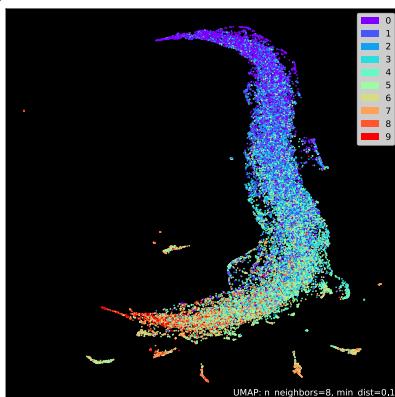
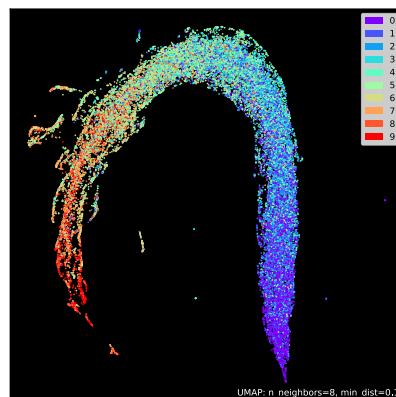
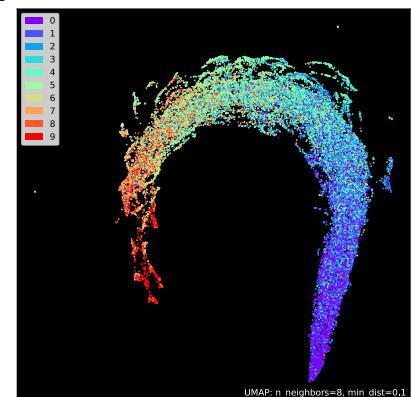
a**b****c****d****e****f**

Table 1. Performance comparison of ConBoTNet and other competing methods

Method	5-CV		LOMO		BD2023 (<i>binary</i>)		BD2023 (IC_{50})		Binding core No. of correct/No. of total
	AUC	PCC	AUC	PCC	AUC	AUC	PCC		
NetMHCIIpan-3.2	0.847	0.679	0.775	0.544	0.571	0.802	0.616	45/51	
NetMHCIIpan-4.0	- ^a	-	-	-	0.577	0.814	0.653	-	
PUFFIN	0.846	0.676	0.768	0.525	-	-	-	-	
DeepSeqPanII	0.741	0.488	0.739	0.485	0.569	0.704	0.416	10/51	
DeepMHCII	0.856	0.690	0.783	0.558	0.620	0.831	0.674	44/51	
ConBoTNet	0.865	0.709	0.789	0.568	0.637	0.845	0.682	45/51	

^a '-' indicates that the corresponding value does not exist.