IMMUNOLOGY   ORIGINAL ARTICLE

# NETMHCSTAB – predicting stability of peptide–MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery

Kasper W. Jørgensen,[1,*] Michael Rasmussen,[2,*] Søren Buus[2] and Morten Nielsen[1,3]

[1]Department of Systems Biology, Centre for Biological Sequence Analysis, Technical University of Denmark, Lyngby, [2]Laboratory of Experimental Immunology, University of Copenhagen, Copenhagen N, Denmark, and [3]Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín, San Martín, Buenos Aires, Argentina

## Summary

Major histocompatibility complex class I (MHC-I) molecules play an essential role in the cellular immune response, presenting peptides to cytotoxic T lymphocytes (CTLs) allowing the immune system to scrutinize ongoing intracellular production of proteins. In the early 1990s, immunogenicity and stability of the peptide–MHC-I (pMHC-I) complex were shown to be correlated. At that time, measuring stability was cumbersome and time consuming and only small data sets were analysed. Here, we investigate this fairly unexplored area on a large scale compared with earlier studies. A recent small-scale study demonstrated that pMHC-I complex stability was a better correlate of CTL immunogenicity than peptide–MHC-I affinity. We here extended this study and analysed a total of 5509 distinct peptide stability measurements covering 10 different HLA class I molecules. Artificial neural networks were used to construct stability predictors capable of predicting the half-life of the pMHC-I complex. These predictors were shown to predict T-cell epitopes and MHC ligands from SYFPEITHI and IEDB to form significantly more stable MHC-I complexes compared with affinity-matched non-epitopes. Combining the stability predictions with a state-of-the-art affinity predictions NETMHCCONS significantly improved the performance for identification of T-cell epitopes and ligands. For the HLA alleles included in the study, we could identify distinct sub-motifs that differentiate between stable and unstable peptide binders and demonstrate that anchor positions in the N-terminal of the binding motif (primarily P2 and P3) play a critical role for the formation of stable pMHC-I complexes. A webserver implementing the method is available at www.cbs.dtu.dk/services/NetMHCstab.

**Keywords:** cytotoxic T lymphocyte epitopes; immunoinformatics; MHC–peptide stability; peptide immunogenicity.

## Introduction

Major histocompatibility complex class I (MHC-I) molecules play a pivotal role in the generation of specific immune responses mediated by cytotoxic T lymphocytes (CTLs). MHC-I molecules sample peptides derived from intracellular proteins, translocate them to the cell surface, and display them to CTLs, allowing immune scrutiny of the ongoing intracellular metabolism leading to the detection of the presence of intracellular pathogens. It has been estimated that only 1 in 200 peptides will bind to a given MHC-I molecule with an affinity stronger than 500 nM.[1] Given this high specificity, binding affinity to MHC-I has naturally been a central focus when developing tools for identification of immunogenic peptides.

Accurate and reliable *in-silico* methods predicting the affinity of peptide binding to MHC-I have been developed over the last decades, supporting with great success the rational discovery of T-cell epitopes, reviewed in refs 2,3. However, other studies[4] have clearly demonstrated that not all peptide binders are necessarily immunogenic indicating that factors other than binding affinity are determinants of peptide immunogenicity. To fulfil the antigen-presenting function, MHC-I molecules must not only bind the peptides generated inside the cell, but also retain them at the cell surface while waiting for the arrival

of extremely rare circulating members of one or more CTL clones of the appropriate specificity. One factor other than affinity that could determine peptide immunogenicity is therefore the stability of the peptide–MHC-I interaction, as complexes with low stability would disassociate before encountering the appropriate CTL clone. The idea of stability being a better predictor than affinity of immunogenicity was proposed.[5–7] In a recent study, Harndahl et al.[8] showed for a set of vaccinia virus peptides binding to HLA-A*02:01, that 30% of the non-immunogenic peptides had (predicted) half-lives for the peptide–MHC-I (pMHC-I) complex below 1 hr, whereas all immunogenic peptides had longer half-lives. Hence, a large proportion of peptides hitherto classified as being non-immunogenic because of 'holes in the T-cell repertoire' were explained in terms of unstable pMHC-I interactions.

Here, we extend the study by Harndahl et al. aiming to demonstrate that the findings for HLA-A*02:01 are generally valid for any MHC-I molecule. Using a high-throughput scintillation proximity assay measuring the half-life of pMHC-I complexes,[9] we generated a large panel of individual pMHC-I stability measurements for 10 prevalent HLA-I molecules covering 8 of the 12 common human MHC-I supertypes.[10,11] Based on such stability measurements, in silico methods were generated for the prediction of half-lives of peptide–MHC-I interactions for the 10 HLA molecules, and the predictive models are used to quantify if immunogenic peptides share a signature in stability different from non-immunogenic binders. Integrating the in silico stability prediction model with state-of-the-art affinity predictions using NetMHCcons,[12] we next evaluated the impact for stability predictions for the rational identification of CTL epitopes.

## Methods

### Artificial neural network training

The data for training of the artificial neural networks were split into five sets in a typical fivefold cross-validation scheme, where four-fifths of the data were for training and the last fifth was for testing and early stopping. This was repeated five times so that all test sets (one of five) were used for evaluation alternately. In this way, the test sets would be independent of the training sets, minimizing the risk of over-fitting the data. Networks were trained as described in Nielsen et al.[13] using either Blosum50 encoding with a normalization factor of 5 or sparse encoded with one of the 20 inputs being 0·95 and the remaining 19 being 0·05. The measured half-life values were transformed from hours to a value falling in the range from 0 to 1. The transformation used was, $s = 2^{-2/Th}$, where $s$ is the transformed value and $Th$ is the half-life measured in hours. This relation was used for all molecules except for

HLA-B*40:01, which had 'unusual' unstable pMHC-I complexes. Here, the relation $s_{B40:01} = 2^{-0.7/Th}$ was used. Using this transformation scheme, a transformed value of 0·5 corresponds to a half-life of 2 hr, except for HLA-B*40:01, where 0·5 corresponds to 0·7 hr.

### Evaluation methods

The Pearson's correlation coefficient was used to evaluate performances of the artificial neural networks. For epitope/ligand data the AUC (area under the receiver operating characteristic curve) was used. When calculating the receiver operating characteristic curves, the source protein was divided into overlapping 9-mers where only the T-cell epitope/ligand was considered positive and all others were considered as negatives. We are aware that when using this definition of epitope/non-epitope some predictions will incorrectly be classified as false positive. However, as the binding motif of MHC class I molecules is very specific, binding only a highly limited repertoire of peptides,[1,14] this misclassified proportion will be very small and will not affect the evaluation in any dramatic manner. Using such receiver operating characteristic curves, the AUC0.1 value corresponding to a specificity of 0·9 was used as a performance measure.[15] Student's paired t-tests were used to evaluate the significance difference between the different methods and approaches used.

### Data

Artificial neural network training. The stability data were generated using the high-throughput scintillation proximity assay, measuring the half-lives of the pMHC-I complexes.[9] In total, the data set consisted of 5509 9-mers (peptides) covering 10 alleles with half-lives measured in hours. The alleles covered are summarized in Table 1.

Seven out of the nine HLA class I supertypes defined by Sette and Sidney[11] were covered – missing supertypes B27 and B58. Of the three additional supertypes (A26, B8

Table 1. Overview of the stability data used for training. Supertype associations are taken from ref. 10

| Allele | No. of peptides | Supertype |
|---|---|---|
| HLA-A*01:01 | 259 | A1 |
| HLA-A*02:01 | 890 | A2 |
| HLA-A*03:01 | 812 | A3 |
| HLA-A*11:01 | 335 | A3 |
| HLA-A*24:02 | 571 | A24 |
| HLA-A*26:01 | 277 | A26 |
| HLA-B*07:02 | 530 | B7 |
| HLA-B*15:01 | 1059 | B62 |
| HLA-B*35:01 | 470 | B7 |
| HLA-B*40:01 | 306 | B44 |
| Total | 5509 | |

and B39) proposed by Lund *et al.*,[10] only A26 was covered. As peptide binding is a prerequisite for measuring stability in the scintillation proximity assay, the stability data set was strongly biased towards strongly binding peptides. As the artificial neural network learning method requires both positive (stable) and negative (unstable) data to perform an optimal training, each data set was enriched with peptides with experimental binding affinities weaker than 20 000 nM[16] (obtained from an in-house database of peptide–MHC affinity measurements). For simplicity, we wanted a universal enrichment size and based on different small-scale pilot studies (results not shown here) 1000 negatives were added to each data set. Each negative peptide was assigned a half-life of 0 hr.

*Benchmarking.* T-cell epitopes and HLA ligands were downloaded from the SYFPEITHI database[17] and the Immune Epitope Database (IEDB).[18] T-cell epitopes and ligands from IEDB that were positive in 'Qualitative Measure' were selected together with all T-cell epitopes and ligands from the SYFPEITHI database meeting the length restriction. Only unique 9-mers were included. First, the SYFPEITHI data were downloaded and hereafter the IEDB data were compared against the SYFPEITHI data and all T-cell epitopes and ligands present in both databases were removed from the IEDB data. The source protein for each T-cell epitope and ligand was compiled using the accession number for each source protein annotated in the IEDB and the 'source of peptides' link provided in the SYFPEITHI database. A substantial subset of the T-cell epitope and ligand data had sequences that did not match the canonical binding motif of the claimed HLA restriction element. In fact, > 15% of the IEDB epitopes were predicted to bind the claimed restriction element with an affinity weaker than 10 000 nM. Given the very high overall accuracy of the state-of-the-art HLA peptide-binding prediction methods, we believe it is very likely that such 'non-binding' peptides are erroneous annotations. Similar observations were also made for data in the SYFPEITHI database not matching the canonical binding motifs.[19] Therefore to focus the analysis on data that, with a very high likelihood, bind the annotated HLA restriction element, epitopes and ligands with a predicted binding affinity weaker than both 500 nM and 2% rank were filtered out.[20] Affinity predictions were calculated using NETMHCCONS.[12] This filter removed approximately 13% of the ligands and 24% of the T-cell epitope data. An overview of the final data sets is given in Table 2.

## Results

### Network performances

Figure 1 shows a bar-plot of the performances of the different network ensembles measured in terms of the

**Table 2.** SYFPEITHI and IEDB data

| Allele | T-cell epitopes SYFPEITHI | Ligands SYFPEITHI | T-cell epitopes IEDB | Ligands IEDB |
|---|---|---|---|---|
| HLA-A*01:01[†] | 13 | 38 | 39[1] | 111 |
| HLA-A*02:01 | 183 | 239 | 606[2] | 84[2] |
| HLA-A*03:01 | 13[3] | 58[3] | 43[3] | 58 |
| HLA-A*11:01 | 11 | 21 | 79[4] | 16[7] |
| HLA-A*24:02 | 38[5] | 23[5] | 78[5] | – |
| HLA-A*26:01 | – | 45 | – | – |
| HLA-B*07:02 | 19[6] | 59[6] | 103[6] | 10 |
| HLA-B*15:01 | 4 | 46 | 10 | 394 |
| HLA-B*35:01 | 10 | – | 40[7] | 1 |
| HLA-B*40:01 | 4 | 25[8] | 7 | – |
| Total | 295 | 554 | 1005 | 674 |

The following serotypes are included in the data sets: [1]HLA-A*01, [2]HLA-A*02, [3]HLA-A*03, [4]HLA-A*11, [5]HLA-A*24, [6]HLA-B*07, [7]HLA-B*35 and [8]HLA-B*40.
[†]Only the serotype HLA-A*01 is present in the SYFPEITHI database.

Pearson's correlation coefficient between predicted and measured half-life times. Three different network ensembles are showed: Blosum[10], Blosum[2,5,10] and Sparse+Blosum. All networks were trained and evaluated using fivefold cross-validation. Blosum[10] indicates networks trained with 10 hidden neurons using Blosum encoding, Blosum[2,5,10] a network ensemble trained with 2, 5 and 10 hidden neurons using Blosum encoding, and finally, Sparse+Blosum is a network-ensemble trained using either sparse or Blosum encoding with 2, 5 and 10 hidden neurons. The combination of sparse and Blosum encoding resulted in the highest performing networks for all data sets. Networks trained on sparse encoding alone were consistently inferior to Blosum encoded networks (results not shown). Network performances ranged from 0·583 (HLA-B*07:02) to 0·815 (HLA-A*11:01). The performance improved for all network ensembles when including more networks in the ensemble. Hence the performance order was: Blosum[10] < Blosum[2,5,10] < Sparse+Blosum, for all data sets used for training.

### Predicted stability of ligands and T-cell epitopes

Figure 2 shows the predicted half-lives of the T-cell epitopes and ligands in the SYFPEITHI data set. The ligands were in general found to form more stable complexes compared with epitopes. Similar results were found for the IEDB data set (see Supplementary material, Fig. S1).

### Affinity-balanced analysis

To investigate if ligands/epitopes were predicted to form more stable complexes compared with other non-epitope/ligand-binding peptides, an affinity-balanced analysis was

**Figure 1.** Bar-plot showing the Pearson's correlation coefficient (PCC) for different network ensembles. All networks were trained and evaluated using fivefold cross-validation. Blosum [10], network trained using Blosum encoded data with 10 hidden neurons. Blosum[2,5,10], network ensemble with 2, 5 and 10 hidden neurons trained using Blosum encoding. Sparse+Blosum, ensemble consisting sparse and Blosum encoded networks with 2, 5 and 10 hidden neurons. Numbers given above bars indicate the PCC for the 'Sparse+Blosum' ensemble.
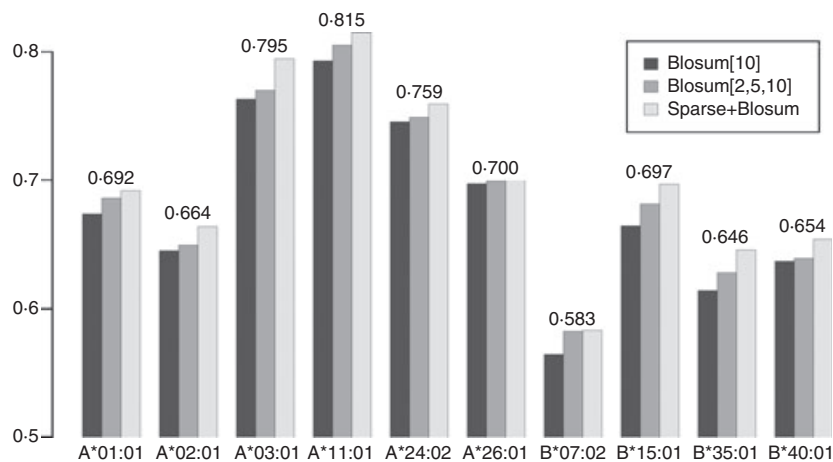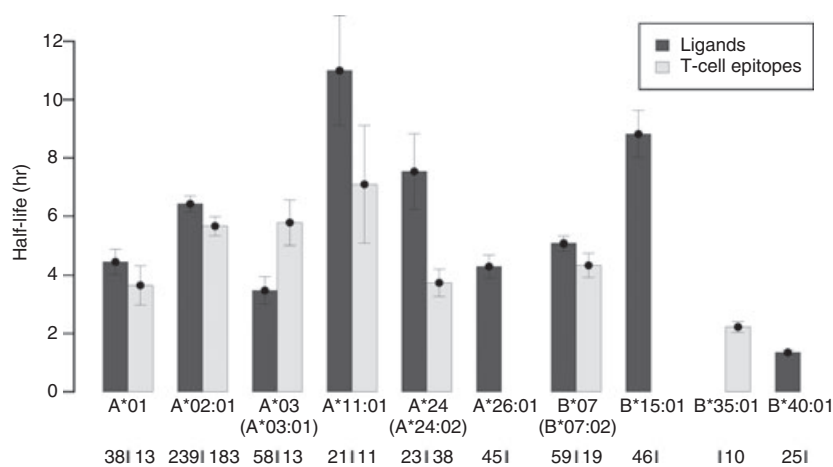
**Figure 2.** Predicted half-life for the SYFPEITHI data. Only data sets with 10 or more peptides are shown. Whiskers indicate standard errors. Alleles in brackets are included in the 'serotype' data set, e.g. the HLA-A*03 data include the data from HLA-A*03:01. Numbers below bars give the number of peptides in the data sets (ligands | T-cell epitopes).

conducted. Using the NETMHCCONS method, binding affinity was predicted for 500 000 random natural 9-mers downloaded from UNIPROT.[21] Each T-cell epitope or ligand was paired with a randomly selected peptide among the 500 000 natural 9-mers with a predicted binding affinity ± 5% of the epitopes/ligands predicted binding affinity. The selection process was balanced, so that approximately 50% of the affinity matched assumed that non-epitopes/ligands had a binding affinity greater than the epitopes/ligands and approximately 50% had a binding affinity weaker than the epitopes/ligands, ensuring no significant difference in affinity between the two groups. The stability was then predicted for the epitopes/ligands and affinity matched non-epitopes/ligands and the difference was tested in a paired Student's $t$-test. Figure 3 shows the results of this comparison for the ligand (Fig. 3a) and T-cell epitope (Fig. 3b) data. The height of the bars indicates the significance level as estimated from the paired Student's $t$-test. The two horizontal dashed lines indicate significance levels of 0·05 (lower line) and 0·01 (upper line).

For most molecules, both ligands and T-cell epitopes were predicted to form more stable complexes than their affinity-matched non-epitope/ligand partner. Three data

sets HLA-A*26:01 (T-cell epitopes), HLA-B*35:01 (ligands) and HLA-B*40:01 (T-cell epitopes) are not included in the figure because < 10 peptides were found for these molecules in both the IEDB and SYFPEITHI databases. For the ligands, 14 of the 15 data sets had the ligands predicted to be more stable than the non-ligands, and in 10 of these, the difference was statistically significant ($P < 0·05$, paired Student's $t$-test). For the T-cell epitopes, 14 of the 15 data sets had the epitopes predicted to be more stable compared with the non-epitopes. However, only in six cases were the differences statistically significant.

## Combining NETMHCCONS and NETMHCSTAB

To investigate to what degree the observation that HLA ligands/epitopes tend to form more stable complexes compared with affinity-matched non-ligands/epitopes could impact a prediction model for epitope/ligand identification; we tested the performance of a simple linear model combining the two properties stability and affinity. Affinity predictions were made using the NET-MHCCONS method, and stability predictions made using the method NETMHCSTAB, developed here. A simple
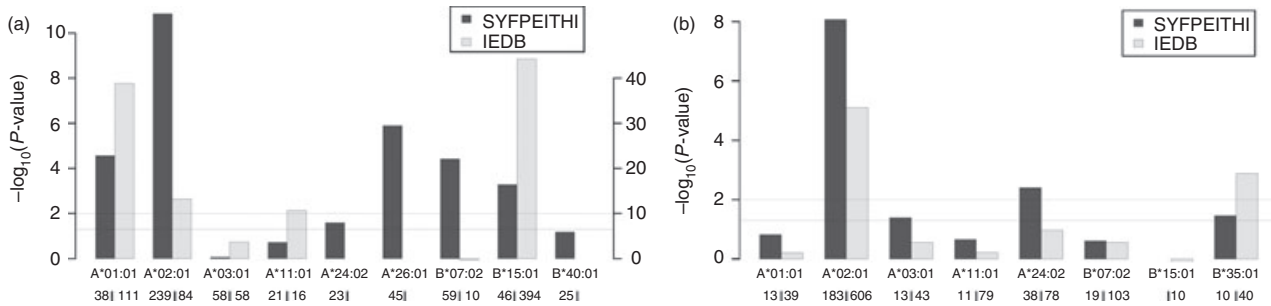
**Figure 3.** Affinity-balanced analysis. (a) Ligands versus non-ligands. (b) T-cell epitopes versus non-T-cell epitopes. The height of the bars gives the $-\log_{10}(P\text{-value})$ as estimated using a Student;s *t*-test. Bars below zero (HLA-B*15:01, T-cell epitopes and HLA-B*07:02, ligands) indicate that the affinity-matched non-epitopes/-ligands were predicted to be more stable than the T-cell epitopes or ligands. The two dashed lines represent $P < 0.05$ and $P < 0.01$, respectively. The numbers below the bars give the number of peptide pairs in the data set. The paired Student's *t*-tests were only performed if more than nine pairs existed. Secondary axis in (a) is for HLA-B*15:01.

weighted sum was used to combine the output from two methods:

$$x = \alpha \cdot NetMHCstab + (1 - \alpha) \cdot NetMHCcons$$

where $x$ is the combined value, $\alpha$ is a value ranging from 0 to 1 and NETMHCSTAB and NETMHCCONS are the output values (between 0 and 1) of the two prediction methods, respectively.

The value of $\alpha$ resulting in the highest performance (average AUC0.1) was estimated in fivefold cross-validation, where weights were optimized on four-fifths of the data and evaluated on the remaining one-fifth. An allele-balanced data set was constructed consisting of a maximum of 50 randomly selected peptides from each allele giving a total of 374 and 355 peptides in the data sets for T-cell epitopes and ligands, respectively. The optimal $\alpha$-value found for each data set was: $\alpha = 0.15$, $\sigma = 0$, where $\alpha$ is the average over the five cross-validations and $\sigma$ is the corresponding standard deviation. The very low variation in optimal $\alpha$ values found in the five cross-validations and the fact that the same optimal value was found for both the T-cell and ligand data sets indicates that the method is highly robust. This is also reflected in the performance of the combined method (see Fig. 4). Here, the performance increase was highly significant when using the combined model compared with NET-MHCCONS alone for both T-cell epitopes and ligands ($P < 0.0001$, in both cases). Also, when evaluated in terms of AUC, the combined model outperformed both of the individual models. The difference was however found to be statistically insignificant ($P = 0.066$ for both T-cell epitopes and ligands).

The remaining 926 T-cell epitopes and 873 ligands not included in the training data used to define the model were used for evaluation of the combined method (see Table 3). Also, in this benchmark the performance of the combined model when measured in terms of AUC0.1, was found to be significantly higher than the NETMHCCONS and NETMHCSTAB methods alone for both
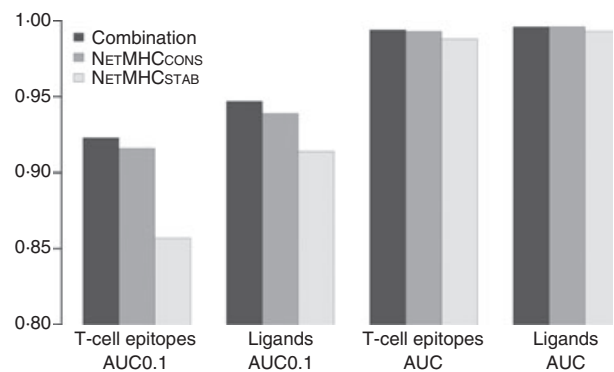


**Figure 4.** Results from fivefold cross-validation. Performance is given as average AUC0.1 and AUC (areas under the receiver operating characteristic curve) for each data set. Data sets used for optimization consisted of 374 and 355 peptides for T-cell epitopes and ligands, respectively. The model parameter $\alpha$ used for the T-cell epitope and ligands data was 0.15.

T-cell epitopes and ligands, ($P = 0.0015$ and $P < 0.0001$, respectively). Likewise, the performance was found, when measured in terms of AUC, to improve when using the combined model. The performance gain however was only significant for ligands ($P < 0.0001$). For T-cell epitopes the *P*-value for the difference was $P = 0.085$.

The gain in AUC values might be hard to translate into actual improvements in the accuracy for T-cell epitope and ligand discovery. Here, a measure like the false-positive proportion might be more useful. We can access this by calculating for each ligand/epitope source-protein pair how many peptides are found with a prediction score greater than the known epitope/ligand. Doing this, we find an average number of false positives of 3.75 and 3.25 (a drop of 15%) for the ligand data set for NETMHCCONS and the combined methods, respectively, and values of 7.75 and 7.50 (a drop of 3%) for the T-cell epitope data set. Using the combined method, taking into account the length of each source protein, these numbers translate

**Table 3.** Results from the evaluation sets. Performance is given as average AUC0.1 and AUC (areas under the receiver operating characteristic curve) for each data set. The data sets 'T-cell epitopes' and 'Ligands' contained 926 and 873 peptides, respectively. The model parameter α used for the T-cell epitope and ligands data was 0·15

| Method | T-cell epitopes (AUC0.1) | Ligands (AUC0.1) | T-cell epitopes (AUC) | Ligands (AUC) |
|---|---|---|---|---|
| Combination | 0·893 | 0·939 | 0·991 | 0·995 |
| NETMHCCONS | 0·888 | 0·926 | 0·991 | 0·994 |
| NETMHCSTAB | 0·848 | 0·922 | 0·986 | 0·994 |

into 98% of the ligands being identified within the top scoring 2·5% of the peptides within the source protein, and for the T-cell epitope data the corresponding value is 91%.

### Motifs and sub-specificities

Finally, motifs and sub-specificities for the different HLA molecules were analysed. We demonstrated earlier for HLA-A*02:01 that stable HLA-binding peptides can be separated from unstable binders in terms of sequence motif characterizing well-defined sub-specificities.[8,22] In particular, for HLA-A*02:01 we could demonstrate that stable binders are distinguished from unstable binders in terms of the motif at the P2 anchor, where stable binders have a very conserved motif compared with the motif of unstable binders. Here, we sought to expand this analysis identifying sub-motifs for each of the 10 HLA molecules separating stable from unstable peptide binders.

Peptide binding to each of the 10 HLA molecules was predicted for a set of 200 000 random natural 9-mer peptides using NETMHCCONS.[12] Binding stability was predicted for the 2000 highest affinity predictions (top 1%) using NETMHCSTAB. Next, for each HLA molecule the peptides were sorted on predicted binding affinity, and subsequently split in a pairwise manner so that the more stable binders were placed in one group and the least stable binders in another group. This setup ensures that the two groups have similar predicted binding affinity and maximal difference in stability. Given this split, we can investigate the differences between stable and unstable binding in a quantitative manner by a direct comparison of the two corresponding binding motifs.

One such direct comparison is shown in Fig. 5 where the sequence logos for stable versus unstable binders for the HLA-A*24:02 and HLA-B*07:02 molecules are shown (sequence logos for stable versus unstable binders for the other eight HLA molecules are included in the Supplementary material, Fig. S2). The average binding affinity of the peptides for the two sub-motifs for HLA-*A24:02 and HLA-B*07:02 are 60 and 94 nM, respectively. In contrast

to this, do the two sub-motifs display highly significant differences in binding stability ($P < 0.001$ in both cases). For HLA-A*24:02 the average stabilities for the two groups are 3·90 hr versus 1·06 hr, and for HLA-B*07:02 the corresponding values are 2·64 hr versus 1·82 hr. For both molecules it is clear that the difference between stable and unstable binders is most pronounced in the N-terminal part of the binding motifs. We can quantify this by comparing the information content in the N-terminal (positions 1–4) and C-terminal (positions 5–9) parts of the binding motifs. The information content at a given position in the binding motif is calculated as the Kullback–Leibler[23] divergence between the amino acid distribution of the peptide binders and a null model defined from the amino acid distribution in a large set of random natural protein sequences. Making this analysis, we find for 8 of the 10 HLA molecules included in this study that the N-terminal information content is higher in the motif for stably bound peptides compared with the motif for unstably bound peptides, whereas no consistent difference was observed in the C-terminal part of the binding motifs (in five cases the stable motifs had higher C-terminal information content, and in five cases the unstable binders had the highest C-terminal information content). Note, that even though the enrichment in N-terminal information content for the motif for stably bound peptides is consistent (8 out of 10 molecules) it is not statistically significant for this small data set ($P = 0.1$, binomial test). The only two molecules not displaying an increased information content in the N-terminal binding motif for stable binders were HLA-A*01:01 and HLA-A*03:01. For HLA-A*01:01, however, a significant difference in the information content at the P3 anchor was observed in the motif for the stable binders (data not shown), hence also suggesting the importance of the N-terminal anchor positions for stabilization of the peptide–HLA complex.

### Discussion

The single most selective step in the MHC class I antigen-processing and presentation pathway is the binding of the peptide to the MHC-I molecule. In the earliest works on characterizing this binding event, significant focus was dedicated to the investigation of both the stability and affinity of the peptide–MHC interactions.[24,25] However, due to the cumbersome and low-throughput nature of the biochemical methods currently used to measure the dissociation of pMHC complexes, the amount of data that has accumulated for the stability of peptide–MHC class I interactions is modest. By way of example, the IEDB contains more than 150 000 distinct peptide-binding affinity measurements to MHC-I molecules with full allelic resolution, whereas the number of stability data is < 3300 (data taken from the IEDB May
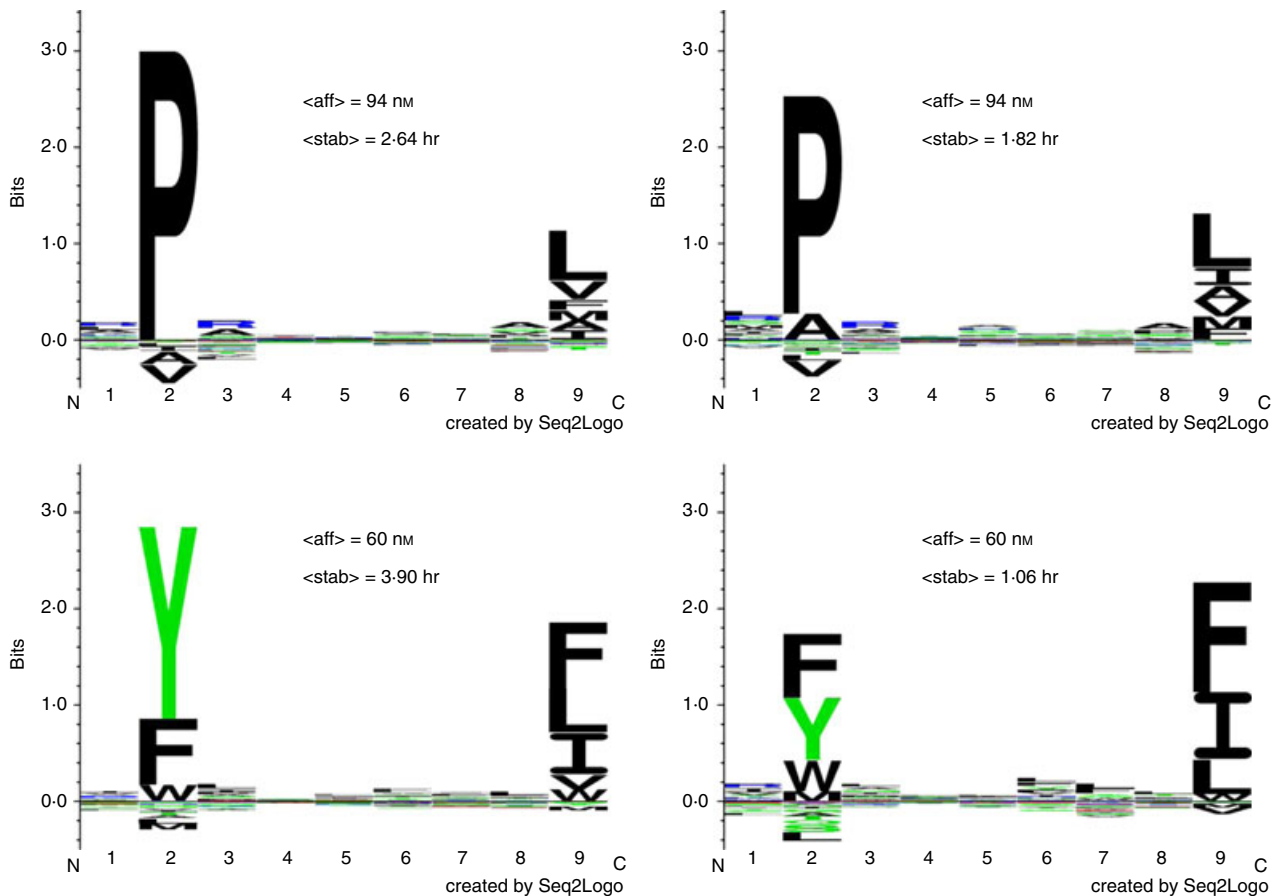
**Figure 5.** Logos for HLA-B*07:02 (upper panel) and HLA-A*24:02 (lower panel). Left panel: Sequence logo representation of the stable binding motifs. Right panel: Sequence logo representation of the unstable binding motifs. The logo-plots were generated using Seq2Logo excluding sequence weighting and pseudo count correction for few observations.[27]

2013). The stability of peptide–MHC-I interactions is therefore a fairly unexplored area and to the best of our knowledge the BIMAS predictor by Parker *et al.*[26] is up to this date the only prediction method available capable of predicting peptide–MHC-I stability. The BIMAS method has not, however, been updated since 1997.

In a recent paper, Harndahl *et al.*[9] proposed a high throughput assay based on a scintillation proximity principle allowing online, real-time monitoring of the dissociation of $^{125}$I-labelled $\beta_2$-microglobulin from recombinant MHC-I heavy chains. In a subsequent publication, the authors demonstrated, using this assay to measure stability of a large set of HLA-A*02:01 binding peptides, that peptide–MHC class I stability was a better predictor than peptide affinity of CTL immunogenicity.[8]

In this study, we extended the Harndahl study to cover 10 HLA molecules (six HLA-A and four HLA-B alleles) covering 8 of the 12 common HLA supertypes. From a large set of stability data, a NetMHCstab method was constructed covering 10 pMHC-I stability predictors based on artificial neural networks. The performance of NetMHCstab measured in a fivefold cross-validation set

up ranged from 0·6 to 0·8 for the 10 networks when evaluated in term of the Pearson's correlation coefficient. Using T-cell epitope and ligands data downloaded from the SYFPEITHI and IEDB databases, the NetMHCstab method was shown to predict both T-cell epitopes and ligands to form very stable peptide–HLA complexes with predicted half-lives > 2 hr. When comparing epitopes and ligands to affinity-matched non-epitopes/non-ligands, we find that epitopes and ligands are predicted to form significantly more stable complexes with the HLA molecules.

Next, we combined affinity (as predicted by NetMHCcons) and stability (as predicted by NetMHCstab) to empower accurate T-cell epitope and ligand predictions. Again based on a large benchmark data set, the combined method integrating both affinity and stability prediction was found to significantly outperform any of the individual methods. The optimal relative weight in the combined method was found to be 15% on stability and 85% on affinity. Hence, NetMHCcons alone has a higher performance than NetMHCstab. However, a direct comparison of the performances of the two methods is not at this

point straightforward. Both methods are constructed in a data-driven manner and NetMHCcons is trained on more than 150 000 data points whereas NetMHCstab is trained only on the 5509 data points described here. It is hence expected that NetMHCcons will achieve the higher performance. Nevertheless, the fact the combined method outperforms the affinity prediction alone clearly demonstrates that stability provides additional information not captured by the affinity predictor, NetMHCcons, and that this information empowers the overall predictive performances.

The gain in predictive performance as measured in terms of the reduction in the number of false predictions when identifying known epitopes/ligands was found to be significant for both T-cell epitopes and HLA ligands when combining stability predictions with prediction of binding affinity. However, in absolute values, the gain for T-cell epitopes was very modest (3%). Many possible explanations for this low gain in predictive performance for T-cell epitopes can be invoked. In our view, one plausible reason is that a large proportion of the T-cell epitope data most likely have been identified using approaches that include (*in silico* or experimental) screening for high HLA affinity, thereby imposing a strong bias on the data that are being entered into the databases; a bias, that is skewed in favour of peptides that match already established affinity motifs of the individual HLA molecules. Investigating this in more detail indeed reveals that > 50% of the references containing 25 or more epitopes in the IEDB applied MHC affinity screening (measured or predicted) to identify epitope candidates. Such a bias is not *a priori* present in the ligand data set, as these data have been derived using mass spectrometry without previous screening for HLA affinity. We therefore believe that as more stability data become available and the accuracy of *in silico* stability predictions improves, this situation will change.

Investigating what discriminates stable from unstable peptide binders, Harndahl et al.[8] found for HLA-A*02:01 that one anchor residue may be sufficient for binding but insufficient for making stable peptide–MHC-I interactions. Here, we extended this analysis and investigated if similar observations could be made for other HLA molecules. Using a large set of natural peptides, we could demonstrate the presence of two binding sub-motifs for all of the HLA molecules included in the study, with one motif corresponding to stable binders and one motif corresponding to unstable binding. Comparing the two sub-motifs for each molecule revealed that especially the presence of amino acids matching the anchor positions in the N-terminal of the binding motif (primarily P2 and P3) was found to have an important role for stable peptide–MHC-I interactions.

In conclusion, we believe that we have demonstrated the significant importance of including pMHC stability predictions in the pathway for rational identification of T-cell epitopes. A webserver implementing the method is available at www.cbs.dtu.dk/services/NetMHCstab.

## Disclosure

The authors declare no conflict of interest.

## References

1 Yewdell JW, Bennink JR. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annu Rev Immunol* 1999; **17**:51–88.

2 Lundegaard C, Lund O, Nielsen M. Predictions versus high-throughput experiments in T-cell epitope discovery: competition or synergy? *Expert Rev Vaccines* 2012; **11**:43–54.

3 Lundegaard C, Lund O, Buus S, Nielsen M. Major histocompatibility complex class I binding predictions as a tool in epitope discovery. *Immunology* 2010; **130**:309–18.

4 Assarsson E, Sidney J, Oseroff C *et al.* A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection. *J Immunol* 2007; **178**:7890–901.

5 Busch DH, Pamer EG. MHC class I/peptide stability: implications for immunodominance, *in vitro* proliferation, and diversity of responding CTL. *J Immunol* 1998; **160**:4441–8.

6 Lazarski CA, Chaves FA, Jenks SA *et al.* The kinetic stability of MHC class II:peptide complexes is a key parameter that dictates immunodominance. *Immunity* 2005; **23**:29–40.

7 van der Burg SH, Visseren MJ, Brandt RM, Kast WM, Melief CJ. Immunogenicity of peptides bound to MHC class I molecules depends on the MHC-peptide complex stability. *J Immunol* 1996; **156**:3308–14.

8 Harndahl M, Rasmussen M, Roder G *et al.* Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. *Eur J Immunol* 2012; **42**:1405–16.

9 Harndahl M, Rasmussen M, Roder G, Buus S. Real-time, high-throughput measurements of peptide-MHC-I dissociation using a scintillation proximity assay. *J Immunol Methods* 2011; **374**:5–12.

10 Lund O, Nielsen M, Kesmir C *et al.* Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics* 2004; **55**:797–810.

11 Sette A, Sidney J. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* 1999; **50**:201–12.

12 Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics* 2011; **64**:177–86.

13 Nielsen M, Lundegaard C, Worning P *et al.* Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* 2003; **12**:1007–17.

14 Rao X, Costa AI, van Baarle D, Kesmir C. A comparative study of HLA binding affinity and ligand diversity: implications for generating immunodominant CD8$^+$ T cell responses. *J Immunol* 2009; **182**:1526–32.

15 Stranzl T, Larsen MV, Lundegaard C, Nielsen M. NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics* 2010; **62**:357–68.

16 Hoof I, Peters B, Sidney J *et al.* NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* 2009; **61**:1–13.

17 Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 1999; **50**:213–9.

18 Vita R, Zarebski L, Greenbaum JA *et al.* The immune epitope database 2.0. *Nucleic Acids Res* 2010; **38**:D854–62.

19 Lamberth K, Roder G, Harndahl M *et al.* The peptide-binding specificity of HLA-A*3001 demonstrates membership of the HLA-A3 supertype. *Immunogenetics* 2008; **60**:633–43.

20 Erup Larsen M, Kloverpris H, Stryhn A *et al.* HLArestrictor – a tool for patient-specific predictions of HLA restriction elements and optimal epitopes within peptides. *Immunogenetics* 2011; **63**:43–55.

21 Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. *Database* 2011; **2011**:bar009.

22 Andreatta M, Lund O, Nielsen M. Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. *Bioinformatics* 2013; **29**:8–14.

23 Kullback S, Leibler RA. On information and sufficiency. *Ann Math Stat* 1951; **22**:76–86.

24 Parker KC, Bednarek MA, Hull LK *et al.* Sequence motifs important for peptide binding to the human MHC class I molecule, HLA-A2. *J Immunol* 1992; **149**:3580–7.

25 Buus S, Sette A, Colon SM, Miles C, Grey HM. The relation between major histocompatibility complex (MHC) restriction and the capacity of Ia to bind immunogenic peptides. *Science* 1987; **235**:1353–8.

26 Parker KC, Bednarek MA, Coligan JE. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J Immunol* 1994; **152**:163–75.

27 Thomsen MC, Nielsen M. Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res* 2012; **40**:W281–7.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** Predicted half-life for the IEDB data.

**Figure S2.** Logos for HLA-A*01:01, HLA-A*02:01, HLA-A*03:11, HLA-A*11:01, HLA-A*26:01, HLA-B*15:01, HLA-B*35:01, and HLA-B*40:01.