The MHC Motif Viewer: A Visualization Tool for MHC Binding Motifs

Nicolas Rapin,[1] Ilka Hoof,[2,3] Ole Lund,[3] and Morten Nielsen[3]

[1]Department of Pharmaceutics and Analytical Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Copenhagen, Denmark
[2]Department of Theoretical Biology/Bioinformatics, Utrecht University, Utrecht, The Netherlands
[3]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark

ABSTRACT

In vertebrates, the onset of cellular immune reactions is controlled by presentation of peptides in complex with major histocompatibility complex (MHC) molecules to T cell receptors. In humans, MHCs are called human leukocyte antigens (HLAs). Different MHC molecules present different subsets of peptides, and knowledge of their binding specificities is important for understanding differences in the immune response between individuals. Algorithms predicting which peptides bind a given MHC molecule have recently been developed with high prediction accuracy. The utility of these algorithms is hampered by the lack of tools for browsing and comparing specificity of these molecules. We have developed a Web server, MHC Motif Viewer, which allows the display of the binding motif for MHC class I proteins for human, chimpanzee, rhesus monkey, mouse, and swine, as well as HLA-DR protein sequences. The binding motif for each MHC molecule is predicted using state-of-the-art, pan-specific peptide-MHC binding-prediction methods, and is visualized as a sequence logo, in a format that allows for a comprehensive interpretation of binding motif anchor positions and amino acid preferences. *Curr. Protoc. Immunol.* 88:18.17.1-18.17.13.

Keywords: MHC • HLA • T cell epitope • binding motif • binding specificity • viewer

INTRODUCTION

In most higher vertebrates, the onset of cellular immune reactions is controlled by the presentation of peptides in complex with major histocompatibility complex (MHC) molecules to T cell receptors (TCR) (Thompson, 1995). The most selective step in the pathway of peptide presentation to the TCR is the binding of the peptide to the MHC molecule (Yewdell and Bennink, 1999). Every human carries 12 classical HLA genes, two alleles of each of the class I loci A, B, and C, and two alleles of each of the class II loci DR, DQ, and DP. Each of these alleles expresses HLA molecules which potentially present a distinct set of antigenic peptides to the immune system (Falk et al., 1991), making all MHC combinations impose a unique signature on the repertoire of peptides presented to the immune system. The human MHC genomic region (called HLA, short for human leukocyte antigen) comprises several thousand allelic variants (Robinson et al., 2001). This immense polymorphism makes rational epitope discovery a daunting task, and has made it highly challenging to correlate immune responses to pathogen infection and host MHC genetic background (Frahm et al., 2007).

For most MHC molecules the binding specificity is still uncharacterized. Of the more than 2000 known HLA class I alleles, for example, the binding specificity has been experimentally characterized for less than 5% (Rammensee et al., 1999; Sette et al., 2005). For nonhuman species, there is an even greater lack of experimental validation. In the case of nonhuman primates, less than 15 alleles have been characterized experimentally (Sette et al., 2005). Characterizing the binding motif of a given MHC molecule requires a significant amount of experimental work. It has been shown that on the order of 100 binding peptides are needed to train an accurate MHC class I binding prediction method. (Yu et al., 2002).

Many MHC molecules share a large fraction of their peptide-binding repertoire, and the use of so-called supertypes has often been considered a solution for dealing with the MHC polymorphism. An MHC supertype is a set of

Current Protocols in Immunology 18.17.1-18.17.13, February 2010
Published online February 2010 in Wiley Interscience (www.interscience.wiley.com).
DOI: 10.1002/0471142735.im1817s88

class I molecules that bind largely overlapping peptide repertoires (Sette and Sidney, 1999; Lund et al., 2004; Sidney et al., 2008). Recent studies, however, seem to indicate that the supertype concept might give a highly oversimplified picture of the diversity in MHC specificity. Many MHC molecules show cross-supertype specificities (Frahm et al., 2007; Perez et al., 2008), and some alleles defined as belonging to the same supertype display very little overlap in peptide repertoire (Hillen et al., 2008).

Development of in silico methods aimed at predicting the binding motif for uncharacterized MHC molecules is therefore important. Several groups have developed prediction methods designed to provide a broad allelic coverage of the MHC polymorphism (Jojic et al., 2006; Nielsen et al., 2007; Hoof et al., 2008; Jacob and Vert, 2008). In contrast to conventional allele-specific methods, these pan-specific methods take both the peptide and the peptide-MHC interaction environment into account, thus allowing for extrapolations to accurately predict the binding specificity of uncharacterized MHC molecules. In the original *NetMHCpan* publication, it was demonstrated that a pan-specific method trained on quantitative human data could predict nonhuman primate binding motifs (Nielsen et al., 2007). Recently, this coverage was extended, and it was demonstrated that a pan-specific method trained on quantitative human, nonhuman primate, and mouse data could accurately predict the binding motifs for HLA-C (Hoof et al., 2008). The *NetMHCpan-2.0* method thus provides quantitative peptide MHC binding predictions for all HLA class I proteins in humans (including HLA-C), as well as chimpanzee (*Pan troglodytes*), rhesus monkey (*Macaca mulatta*), and mouse (*Mus musculus*) MHCs. For MHC class II, Nielsen et al. (2008) have published a method providing peptide-binding predictions covering all HLA-DR alleles with known protein sequence.

While these methods are important for analyzing host-pathogen interactions and for identifying potential T cell epitopes, their usefulness for studying the diversity of the specificity of the immune system within and between species is limited. The authors of this unit have therefore developed a Web interface, the MHC Motif Viewer (*http://www.cbs.dtu.dk/biotools/MHCMotifViewer*), which allows for easy visualization and comparison of predicted binding motifs for MHC class I and class II molecules (Rapin et al., 2008).

Here, we present an updated version of this server with an improved user interface and a novel binding motif visualization format that allows for a comprehensive interpretation of binding motif anchor positions and corresponding amino acid preferences. Further, we explain the use of the Web server for non-expert users and give examples on how the service can be used to interpret complex immunoassay data and understand peptide-MHC binding promiscuity.

METHODS

Pan-Specific MHC Class I and II Prediction Methods

The main engines powering the MHC Motif Viewer are the pan-specific *NetMHCpan* (Nielsen et al., 2007; Hoof et al., 2008) and *NetMHCIIpan* (Nielsen et al., 2008) prediction methods. The *NetMHCpan* method allows for prediction of peptide binding to any MHC class I molecule of known protein sequence. The accuracy of the method has been described in several benchmark studies (Lin et al., 2008a; Zhang et al., 2009). The main difference between the pan-specific *NetMHCpan* method and conventional allele-specific methods like *NetMHC* (Nielsen et al., 2003; Lundegaard et al., 2008), *SMM* (Peters and Sette, 2005), and *ARB* (Bui et al., 2005) lies in the fact that the pan-specific methods can leverage information from multiple MHC molecules to extrapolate the binding specificity for uncharacterized MHC molecules. The *NetMHCpan* method achieves this by including both the peptide amino acid sequence and the amino acids of the MHC molecule defining the binding environment (the so-called pseudo sequence) in the training of the binding-prediction algorithm. This additional information on the binding environment for each peptide binding measurement allows the method to learn peptide-MHC amino acid binding preferences and to extrapolate from these to uncharacterized MHC molecules.

The HLA-DR pan-specific MHC class II binding prediction method, *NetMHCIIpan*, allows for prediction of binding to any HLA-DR molecule of known protein sequence (Nielsen et al., 2008). Like the *NetMHCpan* method, the HLA-DR pan-specific method achieves this by including both peptide (including flanking amino acids) and MHC environment-determining amino acids in the training of the binding-prediction algorithm. This combination of interaction-environment, peptide-core,

and peptide-flanking amino acids allows the *NetMHCIIpan* method to achieve a predictive performance comparable to the state of the art for already characterized MHC class II molecules (Lin et al., 2008b), while simultaneously making it possible to extrapolate and predict the specificity of uncharacterized MHC class II molecules.

Position-Specific Scoring Matrix Construction

To determine the binding motif for each MHC molecule, the binding affinity for a set of 1,000,000 random natural 9-mer peptides (15-mers for the MHC class II binding motifs) was predicted using the *NetMHCpan* method, and the 1% best-binding peptides (1,000) were selected for the position-specific scoring matrix (PSSM) construction. The PSSM was constructed as described by Nielsen et al. (2004) including pseudo count correction for low counts. In short, the PSSM value for amino acid a at position i in the binding motif is calculated as a log-odds score using the relation:

$$S_{ia} = \log \frac{p_{ia}}{q_a}$$

Equation 18.17.1

where p_{ia} is the foreground frequency of the amino acid a at position i, and q_a is the background frequency of amino acid a. The foreground frequency p_{ia} is calculated from the observed amino acid frequency at position i, f_{ia}, combined with the pseudo frequency g_{ia} (Altschul et al., 1997):

$$p_{ia} = \frac{\alpha \cdot f_{ia} + \beta \cdot g_{ia}}{\alpha + \beta}$$

Equation 18.17.2

where $\alpha = N - 1$ (N is the number of peptides) and β is the so-called "weight on prior." To derive the PSSMs for the MHC Motif Viewer, β was set equal to 200 (Nielsen et al., 2004). The pseudo frequency g_{ia} is calculated from the amino acid frequencies f_{ib} at position i using the relation:

$$g_{ia} = \sum_b f_{ib} \cdot q(a|b)$$

Equation 18.17.3

where the sum is over the 20 amino acids, and $q(a|b)$ is the Blosum conditional mutation probability of matching amino acid a to amino acid b (Henikoff and Henikoff, 1992). The background frequencies for the 20 amino acids are obtained from UniProt (UniProt, 2008).

According to the relation used to derive the PSSM values, amino acid a at position i will contribute positively to the binding if its foreground frequency is greater than its background frequency, and, likewise, it will contribute negatively to the binding at position i if its foreground frequency is smaller than its background frequency.

Sequence Logos and How to Interpret Them

Sequence logos (as seen in the left part of Fig. 18.17.1) are a graphical representation of aligned multiple amino or nucleic acid sequences. Sequence logos were originally developed by Tom Schneider and Mike Stephens (Schneider and Stephens, 1990). For each position, the frequency of all 20 amino acids (or 4 bases in the case of nucleic acids) is displayed as a stack of letters. The total height of the stack represents the sequence conservation, while the individual height of the symbols relates to the relative frequency of that particular symbol at that position. This representation is more precise than a simple consensus sequence. The total height is expressed in bits. The higher the stack at a given position, the more conserved the position across all the sequences, and the higher the information content for this position. The total height of the stack, i.e., the information content (R) in bits, is calculated using Claude Shannon's measure (Shannon, 1948) of uncertainty (H) at a given position i:

$$H_i = -\sum_b f_{b,i} \cdot \log_2 f_{b,i}$$

Equation 18.17.4

summing over the twenty amino acids, and where i is the i^{th} position in the protein sequence alignment, and $f_{b,i}$ the frequency of amino acid b at position i. H_i is expressed in bits. The information content R_i at position i is expressed as:

$$R_i = \log_2(20) - H_i = \log_2(20) + \sum_b f_{b,i} \cdot \log_2 f_{b,i}$$

Equation 18.17.5

The relative height of every letter representing a particular amino acid b at position i is proportional to its frequency $f_{b,i}$.

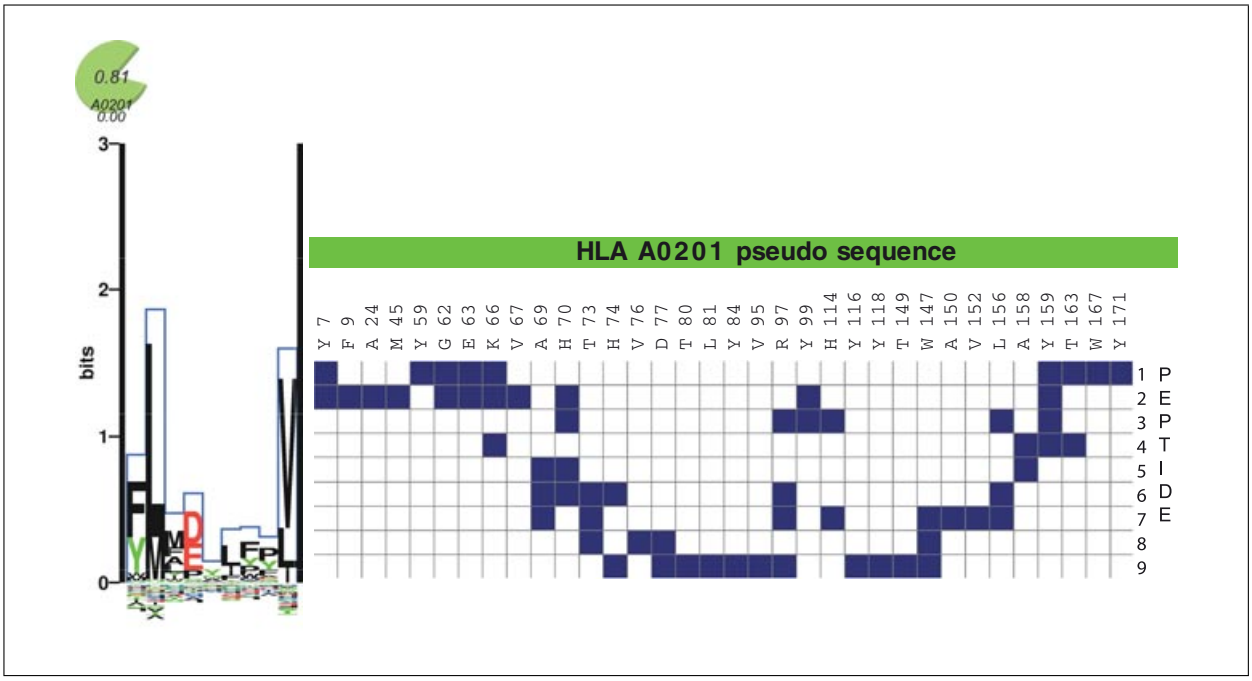

Figure 18.17.1 Left: Kullback-Leibler (KL) sequence logo for the HLA-A*0201 allele. The KL information content is plotted along the 9-mer peptide sequence (solid blue line). Amino acids with positive influence on the binding are plotted on the positive *y* axis, and amino acids with a negative influence on binding are plotted on the negative *y* axis. The relative height of each amino acid is given by Equation 18.17.7, in the text. Right: The contact map for the HLA-A*0201 allele visualizes which residues of the MHC pseudosequence are in contact with which positions in the 9-mer peptide. For color figure go to *http://www.currentprotocols.com/protocol/im1817*.

In the logo plots used in the MHC Motif Viewer Web site, the amino acids are colored according to their physicochemical properties:

Acidic [DE]: red
Basic [HKR]: blue
Hydrophobic [ACFILMPVW]: black
Neutral [GNQSTY]: green.

Modified Kullback-Leibler Logo Representation

In the Shannon information equation, it is assumed that the different symbols (amino acids or nucleic acids) have equal background distribution. In nature, amino acids are found with different frequencies. The Kullback-Leibler (KL) logos, as opposed to the logos described earlier, explicitly take this difference in background amino acid distribution into account when estimating the information content at each position (Kullback and Leibler, 1951). The KL information content is calculated as:

$$R_i = \sum_b f_{b,i} \cdot \log_2 \frac{f_{b,i}}{q_b}$$

Equation 18.17.6

where $f_{b,i}$ is the observed frequency of amino acid i at position b, and q_b is the corresponding background probability. Again, the background probability is derived from large sequence databases like UniProt (UniProt, 2008). Note that for a uniform background frequency distribution, q_b is equal to 1/20 for all twenty amino acids, and the KL information content reduces to the Shannon information content.

For the KL logos of the MHC Motif Viewer, the relative height of amino acid b at position i is proportional to the corresponding term in the relation for the KL information content:

$$R_{b,i} = \frac{f_{b,i} \cdot \log_2 \frac{f_{b,i}}{q_b}}{\sum_c f_{c,i} \cdot \log_2 \frac{f_{c,i}}{q_c}}$$

Equation 18.17.7

In the KL logos, amino acids with a negative log odds value are depicted as upside-down characters on the negative *y* axis, and amino acids with a positive log odds value as upright characters above the positive *y* axis. This way the logo directly reflects the matrix used to generate it and allows for a direct interpretation of which amino acids will have a positive or negative influence on the binding affinity, respectively. A blue histogram is used

to denote the total information content at each position (see left part of Fig. 18.17.1).

Contact Maps

An essential feature of the MHC Motif Viewer interface is the visualization of contact maps. The contact map displays the potential interactions between a peptide sequence and the MHC binding environment (the MHC pseudo sequence). The contact residues in the MHC molecule are defined as being within 4.0 Å of the peptide in any of a representative set of HLA-A and -B protein structures with a bound nonamer peptide. Only residues polymorphic across the HLA-A, B, and C alleles are included giving rise to a pseudo sequence consisting of 34 amino acid residues. Note that due to multiple possible conformations, the central peptide residues could choose to interact with different subsets of residues in the binding groove (Nielsen et al., 2007). An example of such a contact map for the HLA-A*0201 allele is shown in Figure 18.17.1.

Predictive Performance

The predictive performance of the *NetMHCpan* method is to a very high degree determined by the density of characterized MHC class I molecules in the immediate neighborhood of an uncharacterized MHC molecule (Nielsen et al., 2007). This finding was employed in the work by Hoof et al. (2008) to estimate the prediction accuracy of *NetMHCpan* in terms of Pearson's correlation coefficient for any given MHC molecule. In short, the reliability index is estimated from

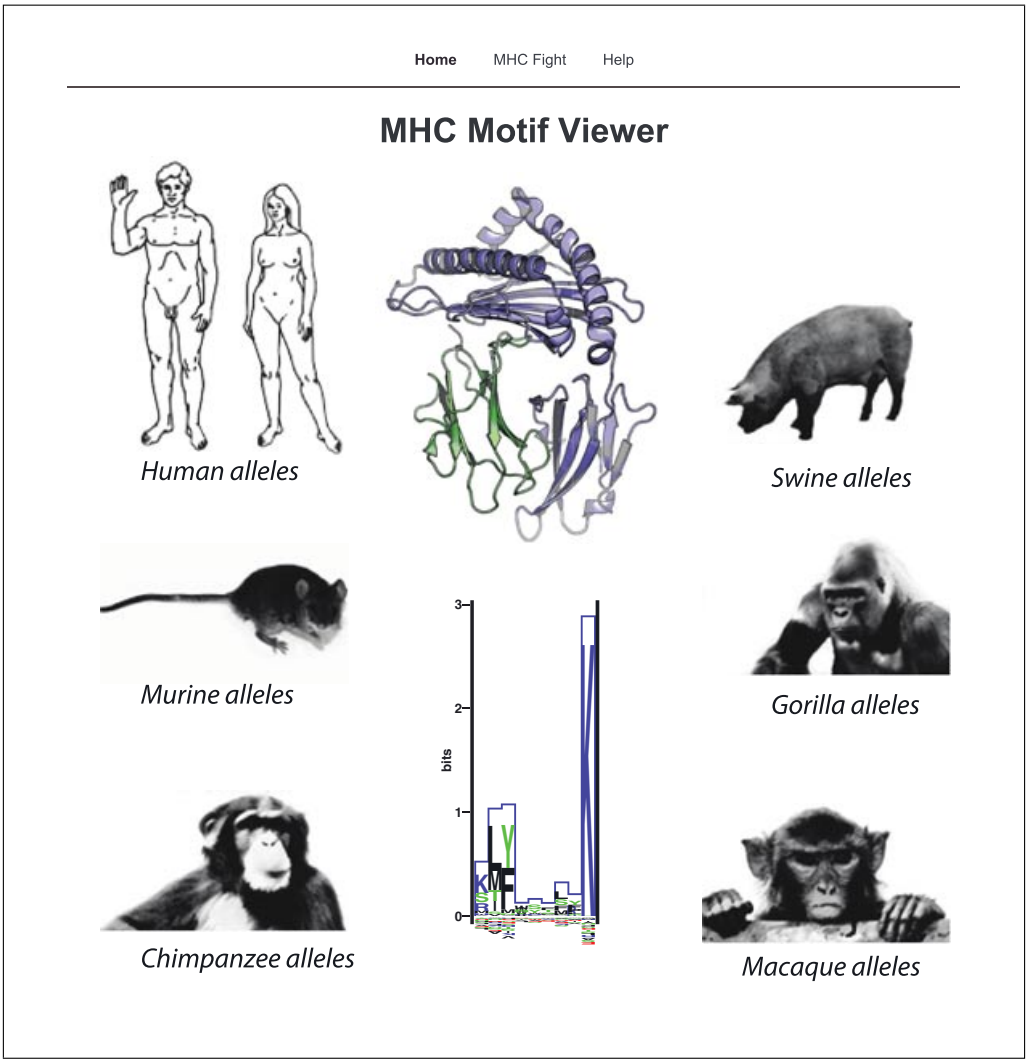

Figure 18.17.2 Home screen for the MHC Motif Viewer Web site. The different pictures are clickable and take the user to different sections of the Web site for the six animal species (human, pig, mouse, gorilla, chimpanzee, and macaque) for which MHC binding motifs were computed. On top of the screen, a quick menu bar allows the user to navigate to the Help and MHC Fight sections.

the pseudo sequence distance to the nearest neighbor for which the binding specificity is well characterized. This measure of prediction accuracy is included in the logo representation of the binding motif as a pie chart indicating the estimated predictive performance (see Fig. 18.17.1). A value of 1 suggests perfect accuracy and a value of 0.0 suggests random predictions.

The MHC Motif Viewer Web Site

The MHC Motif Viewer Web site allows for easy browsing of MHC binding motifs for a large set of alleles. Included HLA alleles cover the class I loci A, B, C, and G and the DR class II families. In addition, the viewer offers binding motifs for mouse, gorilla, macaque, chimpanzee, chicken, and swine MHC class I alleles. The binding motifs are represented by a modified version of the Kullback-Leibler logo, which allows for an easy interpretation of which amino acids are decisive (over-represented) or less important (under-represented) for binding.

Upon loading of the home screen of the Web site, the user is presented with several pictures and a menu on top of the page, which contains links to the MHC Fight program, the home screen, and a help section, as depicted in Figure 18.17.2. The term MHC Fight is inspired by the Google Fight application, where hit counts for two keywords are compared. The MHC Fight application allows for simultaneous visual comparison of the binding motifs of up to four MHC molecules. Six pictures represent the six sections for the different species included in the Web site, namely human, pig, mouse, gorilla, macaque, and chimpanzee. By clicking on the picture for humans, the user is taken to the section regarding human alleles; doing so on the macaque takes the user to the section for macaque, and so on. Some species have been historically more studied (e.g., human) than others (e.g., pig). This is reflected in the amount of data available in the Web site and the organization of the logos for the different species. A summary over the different allele types is given in Table 18.17.1.

Species/Loci Overview

By selecting a species (and, in the case of human, macaque, and chimpanzee, also a locus), the user is redirected to an overview page where several logos for the given species/locus are displayed side by side with the allele name placed below the corresponding logo (Fig. 18.17.3). The page buttons in the upper right part of the page allow for rapid browsing through the binding motifs.

Detailed Allele View

Clicking on an MHC allele displays an enlarged image of the binding motif in question. This page allows the user to view and download the binding motif in terms of a KL logo plot, as well as the corresponding PSSM. The interface, shown in Figure 18.17.4, allows the user to download the motif image in jpg format (logo), the PSSM matrix in Blast profile format (Matrix), and the contact matrix (pseudo-seq) showing which amino acids in the MHC molecule (the pseudo-sequence) are in contact with which residues in the peptide. Next to each MHC class I logo, a reliability index is depicted. The value corresponds to the estimated Pearson correlation coefficient for the *NetMHCpan* predictions for this particular allele. This value is shown together with

Table 18.17.1 Summary Table of the Different Allele Types Represented in the MHC Motif Viewer Web Site[a]

Species	Human	Macaque	Chimpanzee	Pig	Mouse	Gorilla
Allele type	HLA-A (478)	Mamu-A (25)	Patr-A (26)	SLA-10 (9)	H2 (6)	Gogo-B0101 (1)
	HLA-B (792)	Mamu-B (45)	Patr-B (40)	SLA-20 (12)		
	HLA-C (191)			SLA-30 (12)		
	HLA-G (8)					
	DRB1(457)[b]					
	DRB3 (37)[b]					
	DRB4 (7)[b]					
	DRB5 (16)[b]					

[a]The number of alleles of each type is given in parentheses.
[b]Class II alleles.

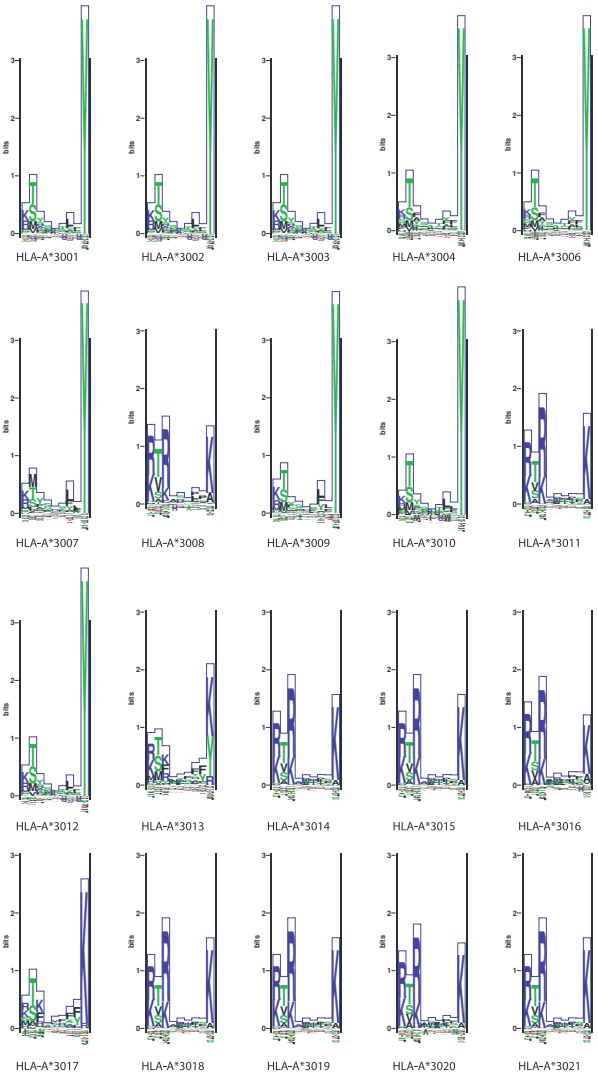

Figure 18.17.3 Species/loci overview. The alleles are arranged on a grid and are clickable. Comparison is made easy because the user has a simultaneous overview of many alleles.

the closest well characterized MHC neighbor and the distance to this neighboring allele. Note that accurate reliability estimations are not available for MHC class II alleles (Nielsen et al., 2008).

APPLICATIONS

The MHC Motif Viewer offers easy browsing of the MHC binding specificity space. Employing the different viewing features, the motif viewer may be used to unravel unexpected similarities between HLA alleles of different serotype as well as unexpected dissimilarities between HLA alleles of the same serotype. In the following, several examples will be presented that demonstrate the value of the MHC Motif Viewer.

Discovering Unexpected Differences in Specificity

The first two digits of each HLA allele name describe its allele family, which is often determined serologically. The full-length protein sequences of the alleles HLA-A*3001 and A*3002 differ only at four positions, corresponding to a sequence identity of 98.9%. However, comparing the binding-motif logos of these two alleles reveals that the binding specificity differs, most notably at the C-terminal anchor position. A*3001 shows a preference for basic amino acids (Lys) at P9, whereas A*3002 prefers the polar amino acid tyrosine (see Fig. 18.17.5A). A look at the contact matrix reveals the reason for this dramatic difference. All four substitutions

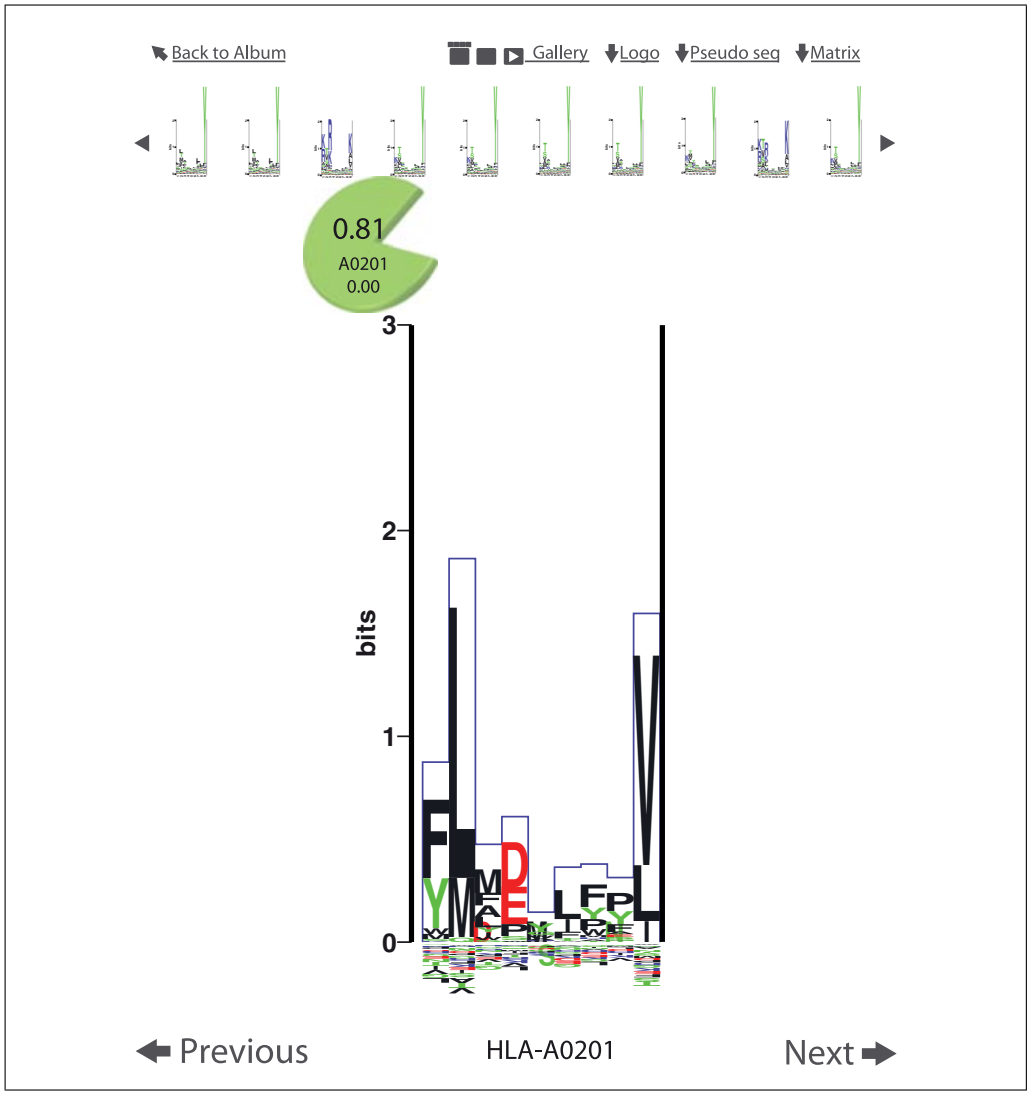

Figure 18.17.4 Detailed view of the human HLA-A*0201 motif logo. The Logo link allows the user to download the motif image in jpg format (logo), the Matrix link allows the user to download the PSSM matrix in Blast profile format, and the Pseudo seq link directs the user to a graphical plot of the contact-matrix (see Fig. 18.17.1). The pie-chart above the logo shows the estimated Pearson correlation coefficient for the *NetMHCpan* predictions for the given allele. This value is shown together with the closest neighbor and the distance to this neighboring allele.

are located in the binding groove of the molecules and are part of the pseudo sequence: Q70H (in contact with peptide positions 2 and 3, Q denotes the amino acid at position 70 in A*3001, H the amino acid in A*3002), V76E (peptide position 8), D77N (peptide positions 8 and 9), and W152R (peptide position 7). Position 77 is a key residue in determining the specificity of the F-pocket (Sidney et al., 2008), and the substitution of the negatively charged Asp (D) by the polar Asn (N) may explain the change of binding specificity from positively charged (Lys) to polar (Tyr).

Another, similar example is given by the HLA-A*6801 and A*6901 alleles. These two molecules differ at five positions on protein sequence level (i.e., 98.6% sequence identity), three of which are part of the pseudosequence (M97R, R114H, D116Y) and take part in the formation of the F-pocket. These substitutions change the C-terminal binding preference from basic (A*6801) to hydrophobic (A*6901); see Figure 18.17.5B.

These two examples illustrate that high similarity on sequence level does not necessarily entail similar binding specificity. A small number of substitutions at key positions of the HLA molecule, namely in the binding pockets that define the specificity of the anchor positions, are sufficient to change the binding specificity dramatically.

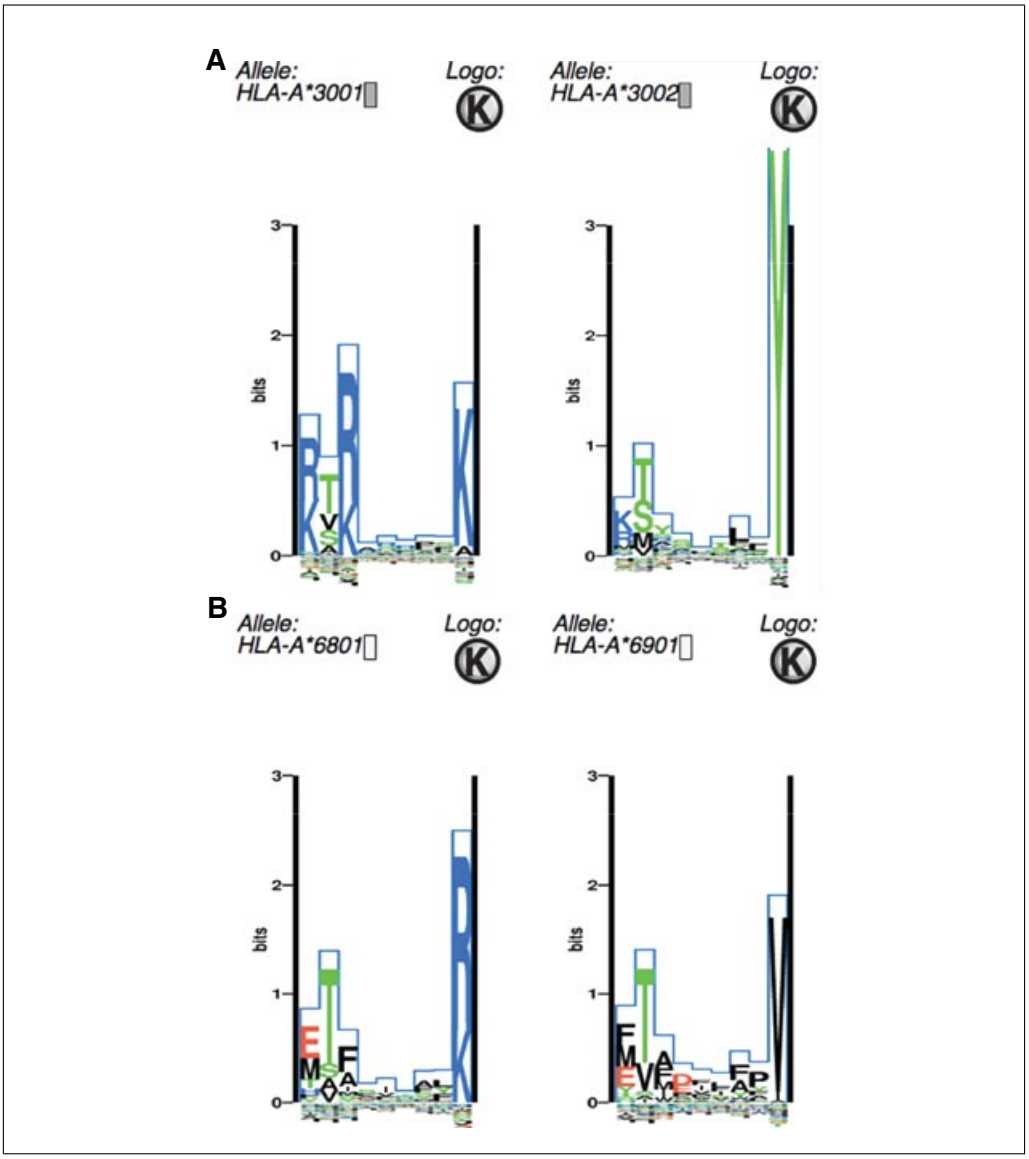

Figure 18.17.5 Motif logos of (**A**) HLA-A*3001 and HLA-A*3002 and (**B**) HLA-A*6801 and HLA-A*6901. Both panels show examples of allele pairs that show a high similarity on a protein-sequence level while revealing clear differences in the C-terminal amino acid preference. Logos were displayed using the MHC Fight Viewer.

Discovering Unexpected Similarities

The HLA-alleles A*0265 and A*0280 have been reported by Sidney et al. (2008) to be of A3-supertype specificity at P9. While our predictions agree with the A*0265 assignment to A3, for A*0280 the predictions suggest that this allele indeed shares the A2-supertype specificity, i.e., hydrophobic preference at P2 and P9 (see Fig. 18.17.6).

Comparing Binding Motifs Across Species

Chimpanzees and humans have been suggested to share peptide-binding motifs (Sidney et al., 2006). Figure 18.17.7 illustrates two such examples: Patr-A*0701 and HLA-A*2402, showing A24 supertype specificity, and Patr-B*1301 and HLA-B*0702, which show B7 supertype specificity.

Finding shared binding specificity for human and chimpanzee MHC molecules may not be very surprising given that chimpanzees are our closest relatives. We can, however, also detect similarities between the binding motifs of human and pig MHC alleles. Figure 18.17.8 illustrates an example of such a pair, SLA-2*0601 and HLA-B*4001, which both show a preference for acidic amino acids at P2 and hydrophobic C-terminal binding preference.

These examples illustrate the possible application of the MHC Motif Viewer in comparing binding motifs across species borders.

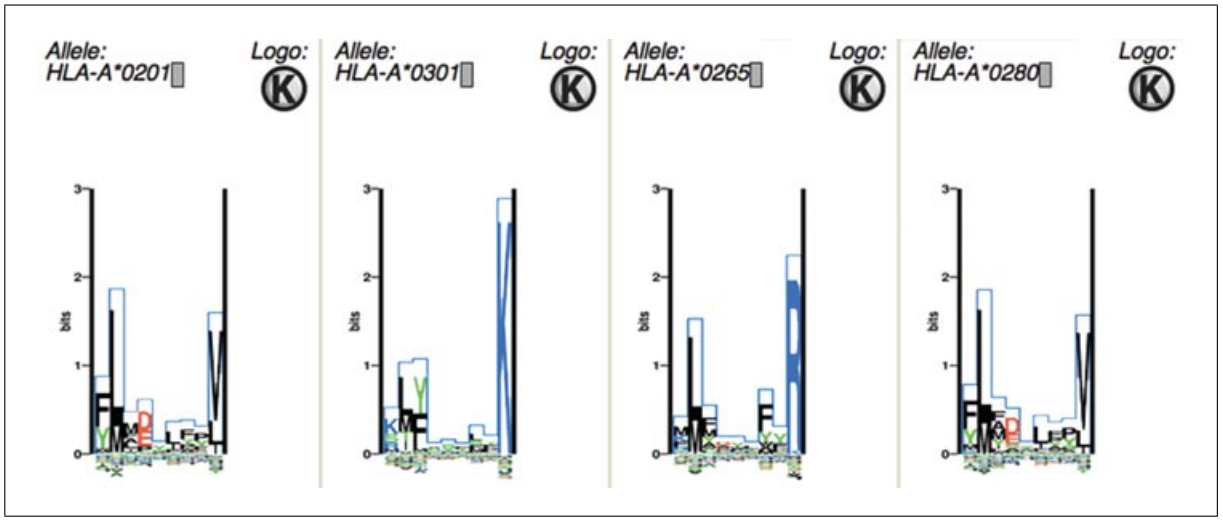

Figure 18.17.6 Motif logos of HLA-A*0201, A*0301, A*0265, and A*0280. The logos reveal the similarity in binding specificity between A*0301 and A*0265, which is unexpected given the serotype of these molecules. A*0280 shares the A2 supertype binding preference, exemplified by A*0201. Logos were displayed using the MHC Fight Viewer.

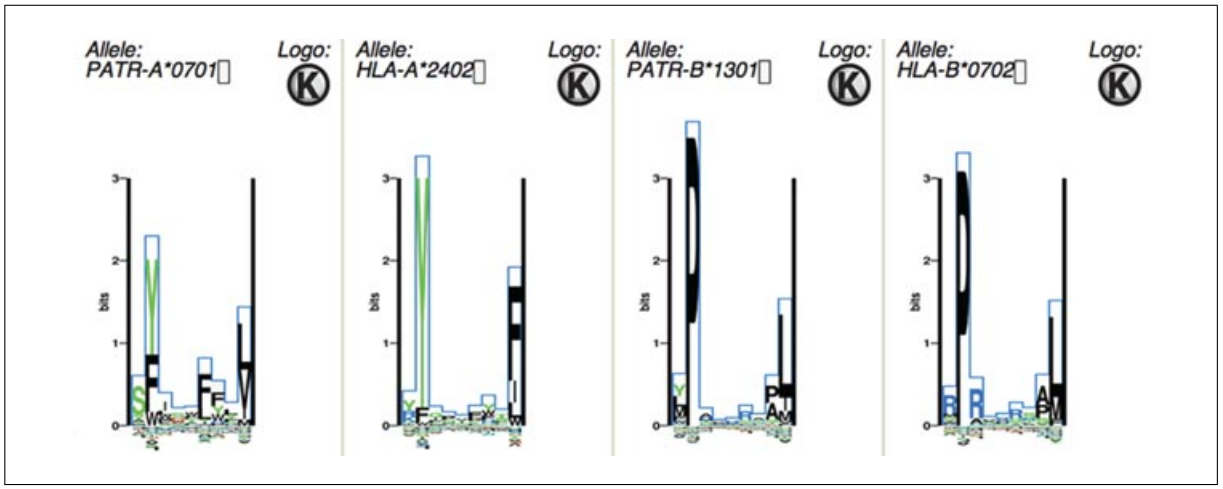

Figure 18.17.7 Motif logos of the Patr-A*0701, HLA-A*2402, Patr-B*1301, and HLA-B*0702. The motif logos illustrate the shared binding specificity that can be observed for some chimpanzee and human MHC class I molecules. Logos were displayed using the MHC Fight Viewer.

Explaining Unexpected CTL Responses

Elispot assays are a common way of screening a patient for epitope-specific CTL responses. Perez et al. (2008) have tested an HIV-1 infected patient cohort for CTL responses against a panel of 9-mer peptides. Some of the observed responses in this study could not be explained based on the patients' HLA genotype, meaning that the peptide that raised the immune response did not fit any of the binding motifs of the responding patient's HLA alleles. Figure 18.17.9 illustrates one such example. A patient, typed to possess the alleles HLA-A*A1101, A*0201, B*4001, and B*3501, showed a high CTL response against the peptide QVPLRPMTY. Surprisingly, none of the expressed HLA molecules has a preference for hydrophobic residues at P2 and polar residues at P9, and the strongest binding affinity predicted using the *NetMHCpan* method (Hoof et al., 2008) is 3000 nM. A closer look at the binding motifs and the peptide, however, reveals that the 9-mer peptide itself contains a nested 8-mer (VPLRPMTY), which fits the binding preference of B*3501 with a predicted binding affinity of 68 nM.

SUMMARY

We have presented the MHC Motif Viewer, which can be used to get a quick impression and overview of the (predicted) binding specificities for a large number of human and non-human MHC molecules.

The motifs are presented as sequence logos in a way that allows for comprehensive

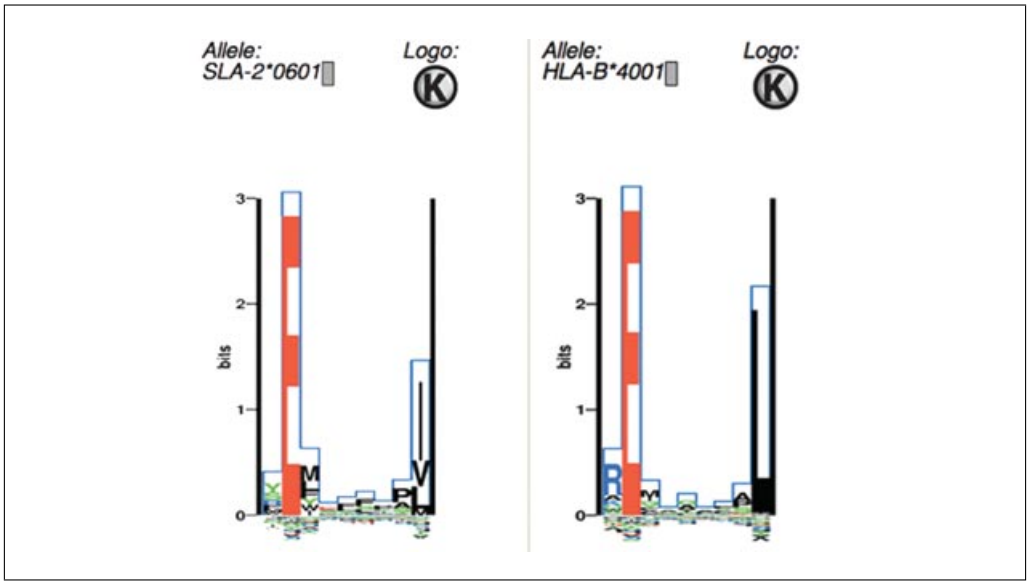

Figure 18.17.8 Motif logos of the pig MHC class I allele SLA-2*0601 and the human allele HLA-B*4001. The logos illustrate their shared binding specificity. Logos were displayed using the MHC Fight Viewer.

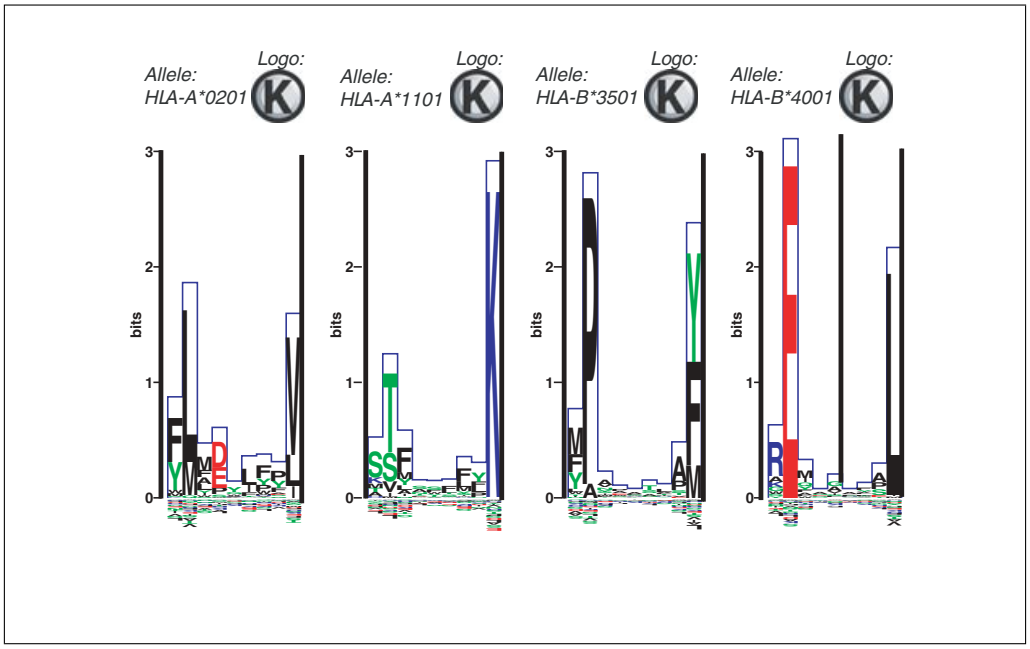

Figure 18.17.9 Motif logos representing a patient's HLA genotype.

interpretation of peptide binding anchor positions and identification of amino acids that promote binding, as well as those that have a negative effect on binding.

For a more detailed comparison, the MHC Fight Viewer enables the user to compare up to four binding motifs side-by-side. This can be used to compare the specificity of different alleles. We have shown how this feature can be used to study unexpected functional *differences* in specificity between alleles that may be serologically identical and genetically similar. We demonstrated how such comparisons might also be useful for the detection of unexpected *similarities* between MHC alleles, and how such similarity might explain peptide cross-reactivity to alleles belonging to different supertypes. The viewer may even be used to study similarities across species borders. We illustrate further how the server can be used to correlate complex immune response data to host MHC genotypes.

The motif viewer is not intended for prediction of MHC-peptide binding. For this, the interested user is referred to the binding-prediction methods *NetMHCpan* and

NetMHCIIpan, which were used to generate the predictions that formed the basis for the binding motifs presented by the MHC Motif Viewer. Both prediction methods are available as online Web servers.

ACKNOWLEDGEMENTS

This work was supported by NIH contracts HHSN266200400083C, HHSN266200400025C, and HHSN266200400006C.

LITERATURE CITED

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.

Bui, H.H., Sidney, J., Peters, B., Sathiamurthy, M., Sinichi, A., Purton, K.A., Mothe, B.R., Chisari, F.V., Watkins, D.I., and Sette, A. 2005. Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. *Immunogenetics* 57:304-314.

Falk, K., Rotzschke, O., Stevanovic, S., Jung, G., and Rammensee, H.G. 1991. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 351:290-296.

Frahm, N., Yusim, K., Suscovich, T.J., Adams, S., Sidney, J., Hraber, P., Hewitt, H.S., Linde, C.H., Kavanagh, D.G., Woodberry, T., Henry, L.M., Faircloth, K., Listgarten, J., Kadie, C., Jojic, N., Sango, K., Brown, N.V., Pae, E., Zaman, M.T., Bihl, F., Khatri, A., John, M., Mallal, S., Marincola, F.M., Walker, B.D., Sette, A., Heckerman, D., Korber, B.T., and Brander, C. 2007. Extensive HLA class I allele promiscuity among viral CTL epitopes. *Eur. J. Immunol.* 37:2419-2433.

Henikoff, S. and Henikoff, J.G. 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.* 89:10915-10919.

Hillen, N., Mester, G., Lemmel, C., Weinzierl, A.O., Muller, M., Wernet, D., Hennenlotter, J., Stenzl, A., Rammensee, H.G., and Stevanovic, S. 2008. Essential differences in ligand presentation and T cell epitope recognition among HLA molecules of the HLA-B44 supertype. *Eur. J. Immunol.* 38:2993-3003.

Hoof, I., Peters, B., Sidney, J., Pedersen, L.E., Sette, A., Lund, O., Buus, S., and Nielsen, M. 2008. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics.* 61:1-13.

Jacob, L. and Vert, J.P. 2008. Efficient peptide-MHC-I binding prediction for alleles with few known binders. *Bioinformatics* 24:358-366.

Jojic, N., Reyes-Gomez, M., Heckerman, D., Kadie, C., and Schueler-Furman, O. 2006. Learning MHC I–peptide binding. *Bioinformatics* 22:E227-E235.

Kullback, S. and Leibler, R.A. 1951. On information and sufficiency. *Ann. Inst. Stat. Math.* 22:76-86.

Lin, H.H., Ray, S., Tongchusak, S., Reinherz, E.L., and Brusic, V. 2008a. Evaluation of MHC class I peptide binding prediction servers: Applications for vaccine research. *BMC Immunol.* 9:8.

Lin, H.H., Zhang, G.L., Tongchusak, S., Reinherz, E.L., and Brusic, V. 2008b. Evaluation of MHC-II peptide binding prediction servers: Applications for vaccine research. *BMC Bioinformatics* 9:S22.

Lund, O., Nielsen, M., Kesmir, C., Petersen, A.G., Lundegaard, C., Worning, P., Sylvester-Hvid, C., Lamberth, K., Roder, G., Justesen, S., Buus, S., and Brunak, S. 2004. Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics* 55:797-810.

Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O., and Nielsen, M. 2008. NetMHC-3.0: Accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Res.* 1:36

Nielsen, M., Lundegaard, C., Worning, P., Lauemoller, S.L., Lamberth, K., Buus, S., Brunak, S., and Lund, O. 2003. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.* 12:1007-1017.

Nielsen, M., Lundegaard, C., Worning, P., Hvid, C.S., Lamberth, K., Buus, S., Brunak, S., and Lund, O. 2004. Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. *Bioinformatics* 20:1388-1397.

Nielsen, M., Lundegaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S., Roder, G., Peters, B., Sette, A., Lund, O., and Buus, S. 2007. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE* 2:E796.

Nielsen, M., Lundegaard, C., Blicher, T., Peters, B., Sette, A., Justesen, S., Buus, S., and Lund, O. 2008. Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS Comput. Biol.* 4:E1000107.

Perez, C.L., Larsen, M.V., Gustafsson, R., Norstrom, M.M., Atlas, A., Nixon, D.F., Nielsen, M., Lund, O., and Karlsson, A.C. 2008. Broadly immunogenic HLA class I supertype-restricted elite CTL epitopes recognized in a diverse population infected with different HIV-1 subtypes. *J. Immunol.* 180:5092-5100.

Peters, B. and Sette, A. 2005. Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics* 6:132.

Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A., and Stevanovic, S. 1999. SYFPEITHI: Database for MHC ligands and peptide motifs. *Immunogenetics* 50:213-219.

Rapin, N., Hoof, I., Lund, O., and Nielsen, M. 2008. MHC motif viewer. *Immunogenetics* 60:759-765.

Robinson, J., Waller, M.J., Parham, P., Bodmer, J.G., and Marsh, S.G.E. 2001. IMGT/HLA Database: A sequence database for the human major histocompatibility complex. *Nucleic Acids Res.* 29:210-213.

Schneider, T.D. and Stephens, R.M. 1990. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* 18:6097-6100.

Sette, A. and Sidney, J. 1999. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and –B polymorphism. *Immunogenetics* 50:201-212.

Sette, A., Fleri, W., Peters, B., Sathiamurthy, M., Bui, H.H., and Wilson, S. 2005. A roadmap for the immunomics of category A-C pathogens. *Immunity* 22:155-161.

Shannon, C.E. 1948. A mathematical theory of communication. *Bell Labs Tech. J.* 27:379-423; 623-656.

Sidney, J., Asabe, S., Peters, B., Purton, K.A., Chung, J., Pencille, T.J., Purcell, R., Walker, C.M., Chisari, F.V. and Sette, A. 2006. Detailed characterization of the peptide binding specificity of five common Patr class I MHC molecules. *Immunogenetics* 58:559-570.

Sidney, J., Peters, B., Frahm, N., Brander, C., and Sette, A. 2008. HLA class I supertypes: A revised and updated classification. *BMC Immunol.* 9:1.

Thompson, C.B. 1995. New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity* 3:531-539.

UniProt. 2008. The universal protein resource (UniProt). *Nucleic Acids Res.* 36:D190-D195.

Yewdell, J.W. and Bennink, J.R. 1999. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annu. Rev. Immunol.* 17:51-88.

Yu, K., Petrovsky, N., Schonbach, C., Koh, J.Y., and Brusic, V. 2002. Methods for prediction of peptide binding to MHC molecules: A comparative study. *Mol. Med.* 8:137-148.

Zhang, H., Lundegaard, C., and Nielsen, M. 2009. Pan-specific MHC class I predictors: A benchmark of HLA class I pan-specific prediction methods. *Bioinformatics* 25:83-89.