

MHC motif viewer

Nicolas Rapin · Ilka Hoof · Ole Lund · Morten Nielsen

Received: 27 May 2008 / Accepted: 8 August 2008 / Published online: 3 September 2008
© Springer-Verlag 2008

Abstract In vertebrates, the major histocompatibility complex (MHC) presents peptides to the immune system. In humans, MHCs are called human leukocyte antigens (HLAs), and some of the loci encoding them are the most polymorphic in the human genome. Different MHC molecules present different subsets of peptides, and knowledge of their binding specificities is important for understanding the differences in the immune response between individuals. Knowledge of motifs may be used to identify epitopes, to understand the MHC restriction of epitopes, and to compare the specificities of different MHC molecules. Algorithms that predict which peptides MHC molecules bind have recently been developed and cover many different alleles, but the utility of these algorithms is hampered by the lack of tools for browsing and comparing the specificity of these molecules. We have, therefore, developed a web server, *MHC motif viewer*, that allows the display of the likely binding motif for all human class I proteins of the loci HLA A, B, C, and E and for MHC class I molecules from chimpanzee (*Pan troglodytes*), rhesus monkey (*Macaca mulatta*), and mouse (*Mus musculus*). Furthermore, it covers all HLA-DR protein sequences. A special viewing feature, *MHC fight*, allows for display of the specificity of two different MHC molecules side by side. We show how the web server can be used to discover and display surprising similarities as well as differences between MHC molecules within and between different species. The *MHC motif viewer* is available at <http://www.cbs.dtu.dk/biotools/MHCMotifViewer/>.

Keywords MHC · HLA · Motifs · Viewer · Class I · Class II

N. Rapin · I. Hoof · O. Lund · M. Nielsen (✉)
Center for Biological Sequence Analysis,
Department of Systems Biology,
Technical University of Denmark,
2800 Lyngby, Denmark
e-mail: mniel@cbs.dtu.dk

Introduction

Presentation of antigenic peptides in complex with major histocompatibility complex (MHC) molecules to T-cell receptors (TCR) controls the onset of cellular immune reaction in the majority of higher vertebrates (Thompson 1995). The most selective step in the presentation of antigenic peptides to TcR is the binding of the peptide to the MHC molecule (Yewdell and Bennink 1999). Each MHC molecule potentially presents a distinct set of antigenic peptides to the immune system (Falk et al. 1991), making rational epitope discovery a daunting task. However, many MHC alleles share a large fraction of their peptide-binding repertoire, and it is often possible to find promiscuous peptides, which bind to a set of different alleles (Lund et al. 2004; Sette and Sidney 1999). Discovery of such peptides is important, since they can be used as vaccines or diagnostic reagents that target a large fraction of the human population. The task of finding epitopes that have the capability of binding to many different MHC molecules can be reduced if MHCs with similar binding specificity can be found.

The human MHC genomic region (called HLA, in short for human leukocyte antigen) comprises more than 3,000 allelic variants (Robinson et al. 2001). Most MHC molecules have uncharacterized binding specificity. Out of the more than 1,500 known HLA class I alleles, for example, less than 5% have their binding specificity characterized experimentally (Rammensee et al. 1999; Sette et al. 2005a). 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. 2005b). The knowledge of MHC motifs is very important in animal models of human disease and for vaccine trials, and characterization of nonhuman MHC binding motifs is, thus, very important for understanding possible implications of these animal models for applications to human vaccines.

Characterizing the binding motif of a given MHC molecule requires a significant amount of experimental work. 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 (Jacob and Vert 2008; Jojic et al. 2006; Nielsen et al. 2007). These methods, in contrast to conventional allele-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, we have extended this coverage and demonstrated that a pan-specific method trained on quantitative human, nonhuman primate, and mouse data can accurately predict the binding motifs for HLA-C and HLA-E loci alleles. The updated *NetMHCpan-2.0* method, thus, provides quantitative peptide MHC binding predictions for all HLA class I proteins in humans (including HLA-C and HLA-E), 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 identifying potential T cell epitopes, their usefulness for studying the diversity of the specificity of the immune system within and between species is limited. We have, therefore, developed a novel web interface that allows for easy visualization and comparison of binding motifs for MHC classes I and II molecules. The motifs were derived using the bioinformatic prediction methods *NetMHCpan* and *NetMHCIIpan*. The validity of the binding motifs was estimated using an independent benchmark data set consisting of 6,533 quantitative peptide binding data points covering 33 human class I alleles, as well as a set of HLA-C and HLA-E ligands.

Results and discussion

Binding motif construction

Binding motifs for MHC class I molecules were estimated using a pan-specific method (Nielsen et al. 2007), *NetMHCpan-2.0*, trained on quantitative human, nonhuman primate, and mouse data. For MHC class II, the binding motifs were estimated using the HLA-DR pan-specific method, *NetMHCIIpan* (Nielsen et al. 2008). For each allele, the top scoring 1% best binders of one million randomly

selected natural peptides were selected and position-specific scoring matrices (PSSM) were calculated using sequence weighting and correction for low counts (Altschul et al. 1997; Nielsen et al. 2004). The binding motifs were visualized using the logo-plot method by Schneider and Stephens (1990). In a sequence logo, the height of a column of letters is equal to the information content at that position, and the height of each letter within a column is proportional to the frequency of the corresponding amino acid at that position. In the logo plots used here, the amino acids are colored according to their physicochemical properties:

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

Binding motif validation

A large set of quantitative peptide MHC binding data were downloaded from the IEDB database (Sette et al. 2005a). The dataset consists of 6,533 peptides and covers 33 HLA-A and HLA-B alleles. The PSSMs were calculated as described above, using the original *NetMHCpan* method trained on human data only. None of the peptides in the evaluation set were included in the training set. For each allele, the predictive performance of the corresponding PSSM was evaluated in terms of the Pearson's correlation between the log-transformed (Nielsen et al. 2003) IC50 value and the summed PSSM prediction score. The result of this validation is shown in Table 1.

From the result shown in Table 1, it can be seen that PSSMs, to a high degree, describe the binding motif of the corresponding HLA alleles, albeit they exhibit an approximately 10% lower Pearson's correlation coefficient than the pan-specific neural network. This is to be expected since the pan-specific method integrates receptor–ligand data representing many different receptors (HLAs) and ligands (peptides), thereby enabling interpolation of information from neighboring HLA molecules. Furthermore, neural networks can take higher order sequence correlations into account.

We have earlier shown how a pan-specific method, trained on quantitative human data, can predict nonhuman primate-binding motifs (Hoof et al. 2008; Nielsen et al. 2007). No quantitative peptide-binding data characterizing the binding motifs of HLA-C and HLA-E alleles are available in the public databases. Limited number of HLA ligand data are available in the SYFPEITHI database for the HLA-C locus (Rammensee et al. 1999) and in the IEDB database for the HLA-E locus (Sette et al. 2005a). These ligand data were used to evaluate the accuracy of the *NetMHCpan* method for HLA-C and HLA-E. For each of the reported ligands, the source protein was identified, and

Table 1 Predictive performance of the PSSMs compared to the *NetMHCpan* method

Allele number	PSSM	NetMHCpan
33	0.557	0.621

The columns give the average performance for 33 HLA-A and HLA-B alleles (6,533 peptide data points) for the PSSM and *NetMHCpan* methods, respectively. The predictive performance was estimated in terms of the Pearson’s correlation coefficient.

the affinity of all nonamers contained in the source protein was predicted. Assuming that all nonamers, with exception of the reported ligand, are nonbinders, the performance was estimated in terms of the relative ranking of the reported ligands among all nonamers contained in the source protein. The average relative ranks obtained were 10.9% and 4.0% for HLA-C and HLA-E, respectively.

The MHC motif viewer

The *MHC motif viewer* provides a visualization of the binding motif of MHC molecules. The interface allows the user to view and download the binding motif in terms of a logo-plot, as well as download the corresponding PSSM. The *MHC motif viewer* interface is shown in Fig. 1.

Fig. 1 *MHC motif viewer* interface. For MHC class I, the viewer includes binding motifs for HLA-A, HLA-B, HLA-C, and HLA-E alleles, as well as chimpanzee, rhesus monkey, and mouse. For MHC class II, the human HLA-DR alleles are included

Clicking on the species or loci of interest displays a table with the corresponding binding motifs. In the table, the allele name is placed below the logo visualization of the binding motif (see Fig. 2a). Clicking on an MHC allele displays an enlarged image of the binding motif (see Fig. 2b) and allows the user to download the motif image in jpg format (logo) or the PSSM matrix in Blast profile format (Matrix). For each allele, there is a link (Pseudo seq) to a table 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 is shown a reliability index. The value corresponds to the estimated Pearson correlation coefficient for neural network predictions on the given allele. This value is shown together with the closest neighbor and the distance to this neighboring allele. The distance was calculated as

[Home](#) [MHC fight](#) [Help](#)

MHC Class I binding motifs



- Human A alleles
- Human B alleles
- Human C alleles
- Human E alleles
- Chimpanzee alleles
- Macaque alleles
- Mouse alleles

MHC Class II binding motifs



- Human DRB1 alleles
- Human DRB3 alleles
- Human DRB4 alleles
- Human DRB5 alleles

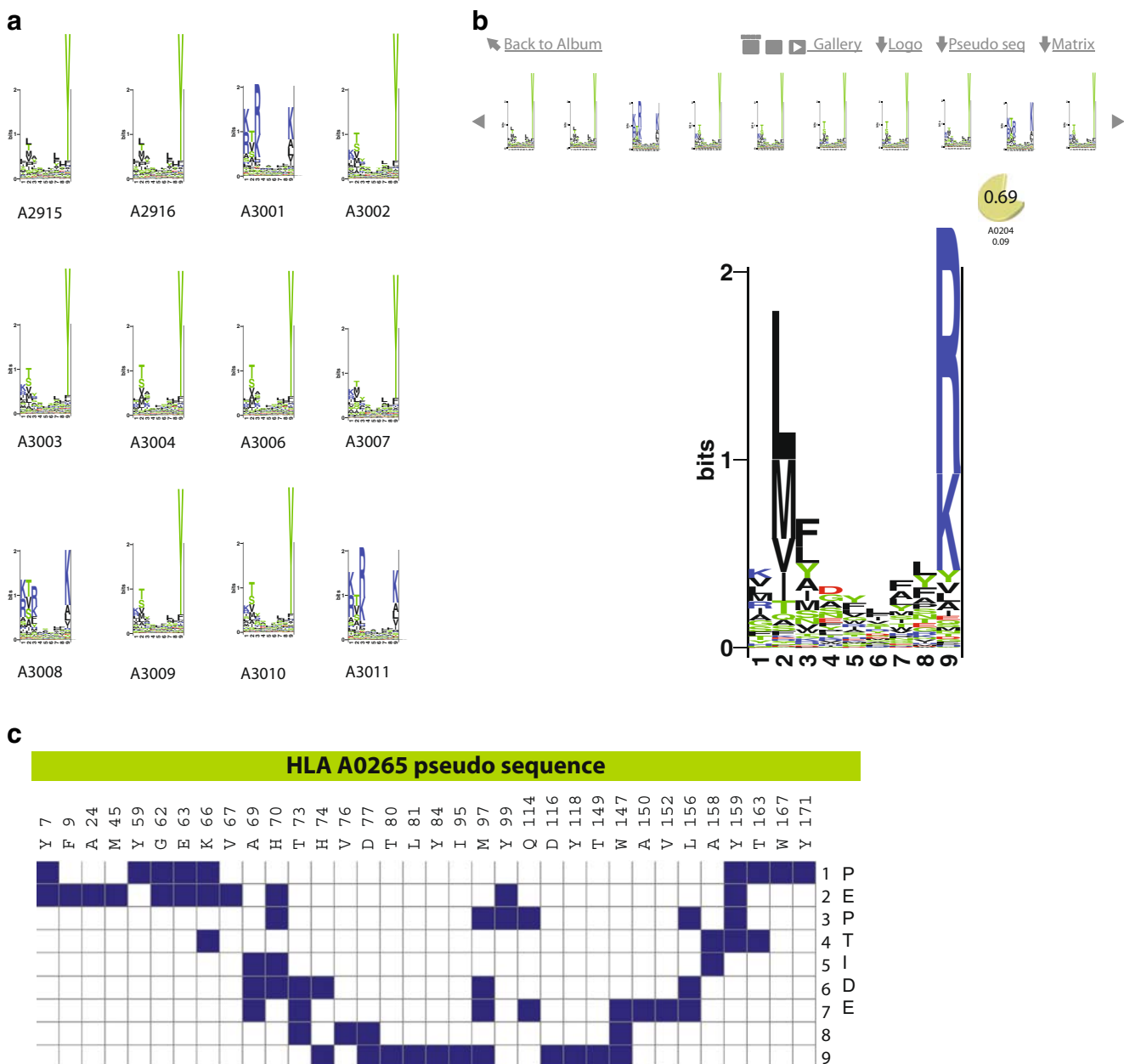


Fig. 2 MHC binding motifs for the HLA-A locus. **a)** displays a motif table, **b)** displays an enlarged image of the predicted HLA-A*0265 binding motif, and **c)** shows the MHC–peptide interaction matrix (displayed by clicking the Pseudo seq link shown in the upper right

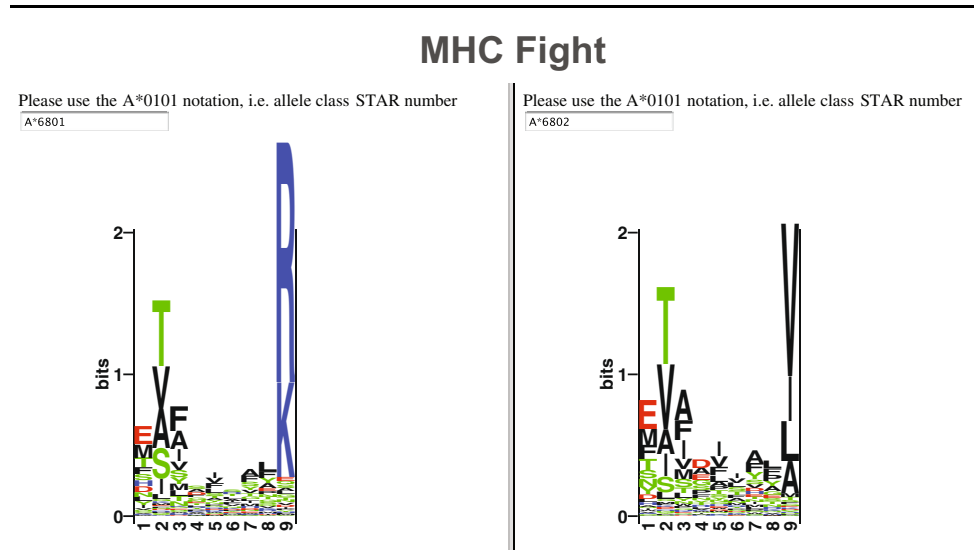
corner in **b)**). The Logo and Matrix links in the upper right corner in **b)** allow for downloads of the binding motif image and PSSM matrix, respectively

described by Nielsen et al. (2007). In short, the reliability index is estimated from the pseudo sequence distance to the nearest neighbor with a well-characterized binding specificity (Nielsen et al. 2007). Note that accurate reliability estimations are not available for MHC class II alleles (Nielsen et al. 2008).

From the motifs displayed in Fig. 2, it is apparent that the HLA-A*0265 allele has a predicted motif that is highly atypical of the A2 serotype. The C terminal binding specificity of this allele shows a preference for basic amino acids, which is a feature of the A3 supertype (Lund et al.

2004). Comparing the pseudo sequence map of the HLA-A*0265 molecule to the typical A2 serotype molecule, HLA-A*0201, we found that these two molecules differ at only four positions in their pseudo sequence. Three of these, 95VI, 97RM, and 116YD, are in contact with the peptide P9 position. Here, the first amino acid is found in the HLA-A*0201 molecule and the second in the HLA-A*0265 molecule. In particular, the last two amino acid substitutions strongly support the finding that the C terminal binding specificity of the HLA-A*0265 allele has a preference for

Fig. 3 *MHC fight viewer*. Comparing the binding motif for the HLA-A*6801 and HLA-A*6802 molecules. The HLA-A*6801 has basic A3 supertype P9 preferences and HLA-A*6802 has hydrophobic A2 supertype preferences



basic amino acids. Similar serotype mixings can be observed for other alleles like HLA-A*0318 (A1 preference at P2), HLA-A*0319 (A2 preference at P9), HLA-A*2914 (A3 preference at P9), and HLA-A*7410 (A1 preference at P9). Another observation is a large degree of specificity mixing within serotypes. The B35 serotype is, for instance, highly mixed with alleles showing similarity to the B7 supertype (proline at P2) and other B35 serotype alleles showing similarity to the B44 supertype (acidic amino acids at P2).

Limited data exist characterizing the MHC binding motif of nonhuman primates. Here, we include motifs for 66 chimpanzee (Patr) and 70 rhesus monkey (Mamu) MHC class I sequences. The Patr protein sequences were obtained from the IPD-MHC sequence database (Ellis et al. 2006) and the Mamu sequences from the NCBI Entrez Protein database. In addition, the *MHC motif viewer* includes motifs

for six mouse MHC class I alleles. Finally, the *MHC motif viewer* offers predicted motifs for the HLA-DR class II loci.

The MHC fight viewer

The *MHC fight viewer* (the link in Fig. 1, top row) allows for direct comparison of the binding motifs of two MHC molecules. The use of the *MHC fight viewer* is illustrated in Fig. 3. Here, the binding motifs of the two alleles HLA-A*6801 and HLA-A*6802 are compared. The figure confirms the known C terminal binding preference of the two alleles, with HLA-A*6801 having A3 supertype preferences, and HLA-A*6802 having A2 supertype preferences (Lund et al. 2004).

The *MHC fight viewer* can also be used to illustrate binding motif similarities between MHC molecules.

Fig. 4 *MHC fight viewer*. Comparing the binding motif for the HLA-A*2402 and chimpanzee Patr-A*0701 molecules

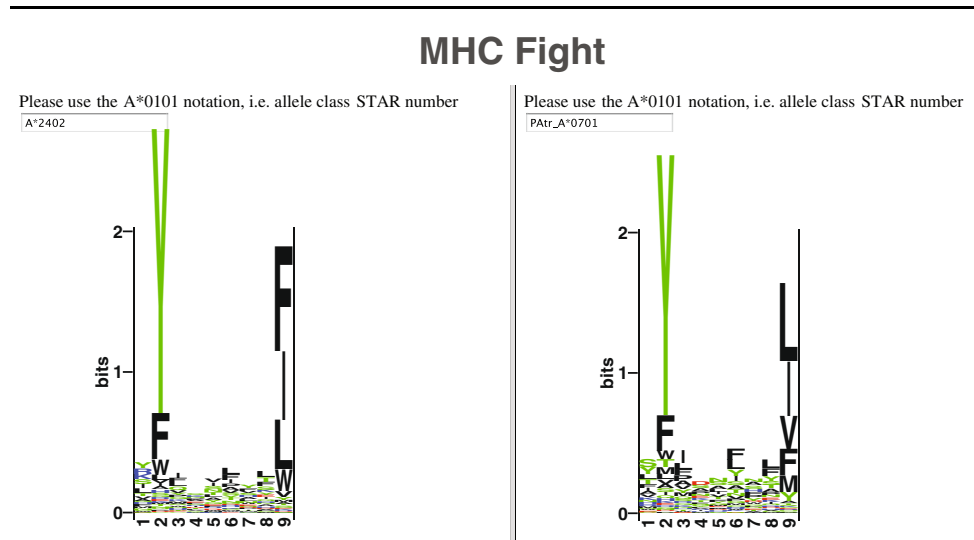


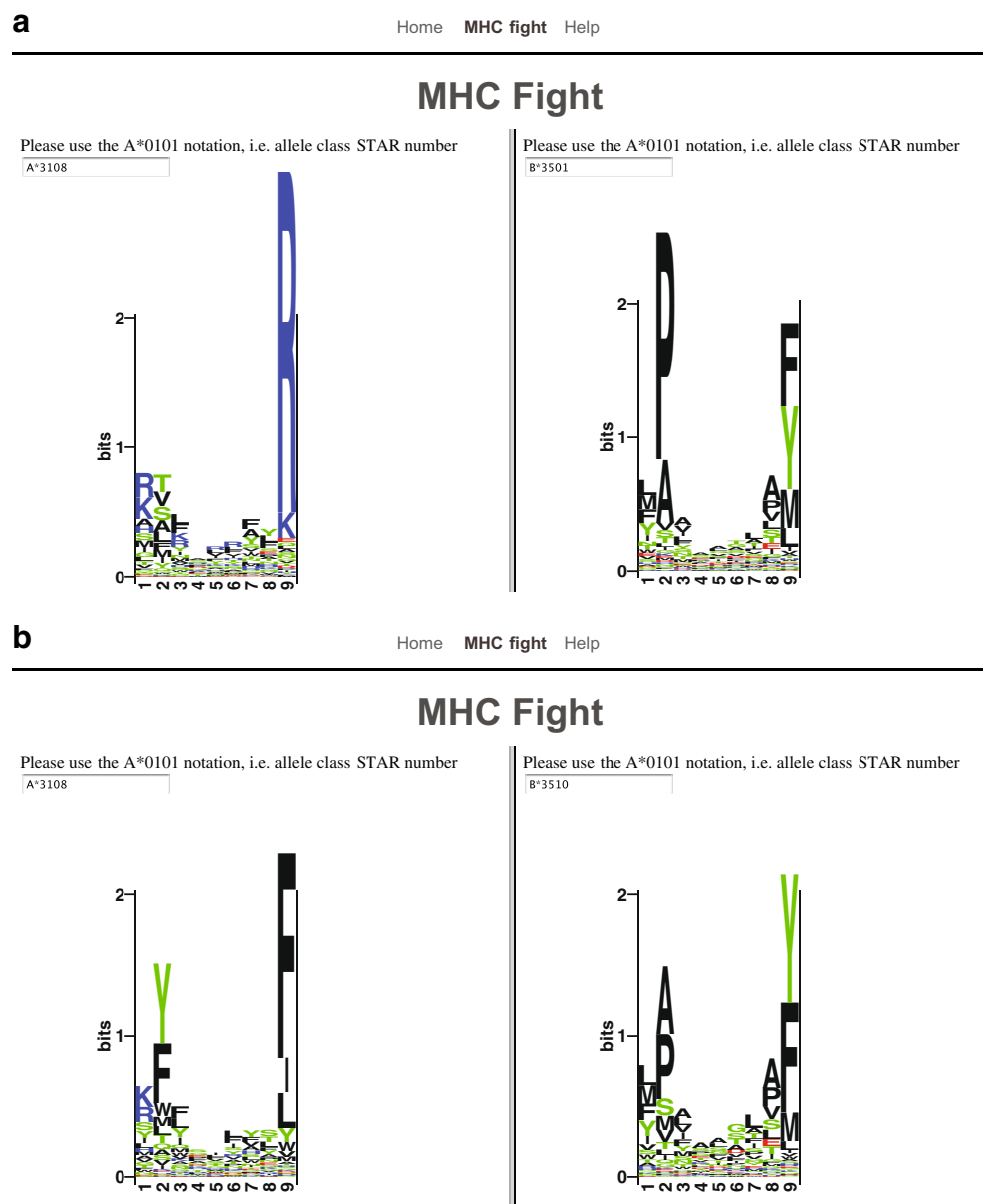
Figure 4 illustrates such a comparison. Here, the binding motif of HLA-A*2402 is compared to the chimpanzee Patr-A*0701 allele. The figure clearly demonstrates the strong similarity between the binding motifs of the two alleles, confirming the observation earlier put forward by Sidney et al. (2006).

The diversity of the MHC molecule world requires tools aimed at biologists and immunologists that focus on ease of use and simplicity. The ability to compare different MHC specificities has applications ranging from experimental design of peptide-binding assays, antigenicity assays, to personal medicine, on a large scale. The use of the *MHC fight viewer* is, thus, a very useful tool for making pairwise comparisons of MHC binding motifs, enabling direct aid in interpreting for instance epitope screening data in patient

cohorts with HLA diversity. Furthermore, it can be used to select peptides that are immunogenic in humans as well as in model organisms.

An example of the use of the *MHC fight viewer* is illustrated in Fig. 5. Frahm et al. (2007) did an analysis of CTL epitope recognition in a cohort of patients with broad HLA diversity and identified a large set of peptides that showed cross-reactivity to HLA molecules of seemingly very different binding specificity. An example of such a peptide is AVLLHEESM from EBV that was found to cross-react to the HLA-A*31 and HLA-B*35 HLA allele. We can try to explain this cross-reactivity using the *MHC fight viewer*. The upper panel of Fig. 5 shows the binding motifs of the HLA-A*3101 and HLA-B*3501 alleles, and clearly, the motifs are very different. However, the lower

Fig. 5 Explaining MHC cross-reactivity. The EBV peptide AVLLHEESM cross-reacts with the HLA-A*31 and HLA-B*35 serotypes. **a)** compares the motifs of HLA-A*3101 to HLA-B*3501. **b)** compares the motifs of HLA-A*3108 to HLA-B*3510



panel of Fig. 5 shows a clear overlap in binding motifs of the HLA-A*3108 and HLA-B*3510 molecules, suggesting a possible explanation for the observed cross-reactivity between the two serotypes.

Conclusion

In conclusion, we believe the *MHC motif viewer* provides a highly useful tool for immunologists to gain insights to the peptide amino acid preference of MHC molecules and, thereby, aiding researchers interpreting their experimental observations.

Acknowledgements This work was supported by the EC contract FP6-2004-IST-4, No.028069 (ImmunoGrid) and NIAID Contracts no. HHSN266200400083C and HHSN26620040006C

References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. doi:10.1093/nar/25.17.3389
- Ellis SA, Bontrop RE, Antczak DF, Ballingall K, Davies CJ, Kaufman J et al (2006) ISAG/IUIS-VIC Comparative MHC Nomenclature Committee report, 2005. *Immunogenetics* 57:953–958. doi:10.1007/s00251-005-0071-4
- Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 351:290–296. doi:10.1038/351290a0
- Frahm N, Yusim K, Suscovich TJ, Adams S, Sidney J, Hrabec P et al (2007) Extensive HLA class I allele promiscuity among viral CTL epitopes. *Eur J Immunol* 37:2419–2433. doi:10.1002/eji.200737365
- Hoof I, Kesmir C, Lund O, Nielsen M (2008) Humans with chimpanzee-like major histocompatibility complex-specificities control HIV-1 infection. *AIDS* 22:1299–1303
- Jacob L, Vert JP (2008) Efficient peptide-MHC-I binding prediction for alleles with few known binders. *Bioinformatics* 24:358–366. doi:10.1093/bioinformatics/btm611
- Jojic N, Reyes-Gomez M, Heckerman D, Kadie C, Schueler-Furman O (2006) Learning MHC I-peptide binding. *Bioinformatics* 22:227–e235. doi:10.1093/bioinformatics/btl255
- Lund O, Nielsen M, Kesmir C, Petersen AG, Lundegaard C, Worning P et al (2004) Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics* 55:797–810. doi:10.1007/s00251-004-0647-4
- Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S et al (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* 12:1007–1017. doi:10.1110/ps.0239403
- Nielsen M, Lundegaard C, Worning P, Hvid CS, Lamberth K, Buus S et al (2004) Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. *Bioinformatics* 20:1388–1397. doi:10.1093/bioinformatics/bth100
- Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S et al (2007) NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS One* 2:796. doi:10.1371/journal.pone.0000796
- Nielsen M, Lundegaard C, Blicher T, Peters B, Sette A, Justesen S et al (2008) Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLOS Comput Biol* 4(7):e1000107
- Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50:213–219. doi:10.1007/s002510050595
- Robinson J, Waller MJ, Parham P, Bodmer JG, Marsh SGE (2001) IMGT/HLA Database - a sequence database for the human major histocompatibility complex. *Nucleic Acids Res* 29:210–213. doi:10.1093/nar/29.1.210
- Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 18:6097–6100. doi:10.1093/nar/18.20.6097
- Sette A, Sidney J (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* 50:201–212. doi:10.1007/s002510050594
- Sette A, Fleri W, Peters B, Sathiamurthy M, Bui HH, Wilson S (2005a) A roadmap for the immunomics of category A-C pathogens. *Immunity* 22:155–161. doi:10.1016/j.immuni.2005.01.009
- Sette A, Sidney J, Bui HH, del Guercio MF, Alexander J, Loffredo J et al (2005b) Characterization of the peptide-binding specificity of Mamu-A*11 results in the identification of SIV-derived epitopes and interspecies cross-reactivity. *Immunogenetics* 57:53–68. doi:10.1007/s00251-004-0749-z
- Sidney J, Asabe S, Peters B, Purton KA, Chung J, Pencille TJ et al (2006) Detailed characterization of the peptide binding specificity of five common Patr class I MHC molecules. *Immunogenetics* 58:559–570. doi:10.1007/s00251-006-0131-4
- Thompson CB (1995) New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity* 3:531–539. doi:10.1016/1074-7613(95)90124-8
- Yewdell JW, Bennink JR (1999) Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annu Rev Immunol* 17:51–88. doi:10.1146/annurev.immunol.17.1.51