Immunogenetics (1999) 50:213–219

REVIEW

Hans-Georg Rammensee · Jutta Bachmann
Niels Philipp Nikolaus Emmerich
Oskar Alexander Bachor · Stefan Stevanović

SYFPEITHI: database for MHC ligands and peptide motifs

Abstract The first version of the major histocompatibility complex (MHC) databank SYFPEITHI: database for MHC ligands and peptide motifs, is now available to the general public. It contains a collection of MHC class I and class II ligands and peptide motifs of humans and other species, such as apes, cattle, chicken, and mouse, for example, and is continuously updated. All motifs currently available are accessible as individual entries. Searches for MHC alleles, MHC motifs, natural ligands, T-cell epitopes, source proteins/organisms and references are possible. Hyperlinks to the EMBL and PubMed databases are included. In addition, ligand predictions are available for a number of MHC allelic products. The database content is restricted to published data only.

Key words MHC · Peptide motif · T-cell epitope

Introduction

The function of MHC molecules is the transfer of information about the current stock of proteins within a cell to the cell surface, thus enabling the immune system to react if necessary, for example by inducing cytotoxic T lymphocytes to kill virus-infected cells or by activating B cells by a helper T lymphocyte. In this context, the peptide specificity of MHC class I and class II molecules is important for the selection of relevant peptides.

The first collection of MHC ligands and peptide motifs was published as the First Listing in the anniversary issue of Immunogenetics in 1995 (Rammensee et al. 1995) and already contained at this time a few hundred entries, mainly consisting of human ligands and T-cell epitopes. By 1997, an enormous amount of information on the peptides associated with MHC molecules had been accumulated and an update of the first collection was published (Rammensee et al. 1997). This collection contained not only MHC peptide motifs, MHC ligands, and T-cell epitopes but also the amino acid sequences of MHC molecules, to enable the elucidation of the structural basis of MHC motifs by analyzing the nature of the pockets involved in the binding process. Since then, even more MHC motifs have become available. In order to meet the needs of colleagues involved with MHC-associated peptides and to provide a handy source of MHC ligands and epitopes and related entries, the database SYFPEITHI was designed and published via the World-Wide-Web (WWW). The name SYFPEITHI was chosen to acknowledge the first MHC-eluted peptide that was directly sequenced (Falk et al. 1991). Data related to peptides eluted from MHC molecules have been compiled in the database; at present, it comprises approximately 2000 entries reported for human, mouse, rat, ape, cattle, and chicken MHC alleles. All entries are linked to the respective sequences of the EMBL database and the abstracts published in PubMed. The possibility of epitope predictions for a limited number of MHC motifs has been opened, and in the near future the molecular mass of the peptides will complement the collection.

H.-G. Rammensee (✉) · J. Bachmann · N.P.N. Emmerich
S. Stevanović
Universität Tübingen, Interfakultäres Institut für Zellbiologie,
Abteilung Immunologie, Auf der Morgenstelle 15,
D-72076 Tübingen, Germany
e-mail: rammensee@uni-tuebingen.de
Tel.: +49-7071-2987628
Fax: +49-7071-295653

O.A. Bachor
EMS Medicinal and Scientific Dataprocessing,
Willy-Brandt-Platz 6, D-68161 Mannheim, Germany
EMS.Mannheim@t-online.de

Materials and methods

All data are stored centrally in a relational client-server database system (RDBMS). The main table of the RDBMS contains examples of ligands and T-cell epitopes, as well as additional informa-

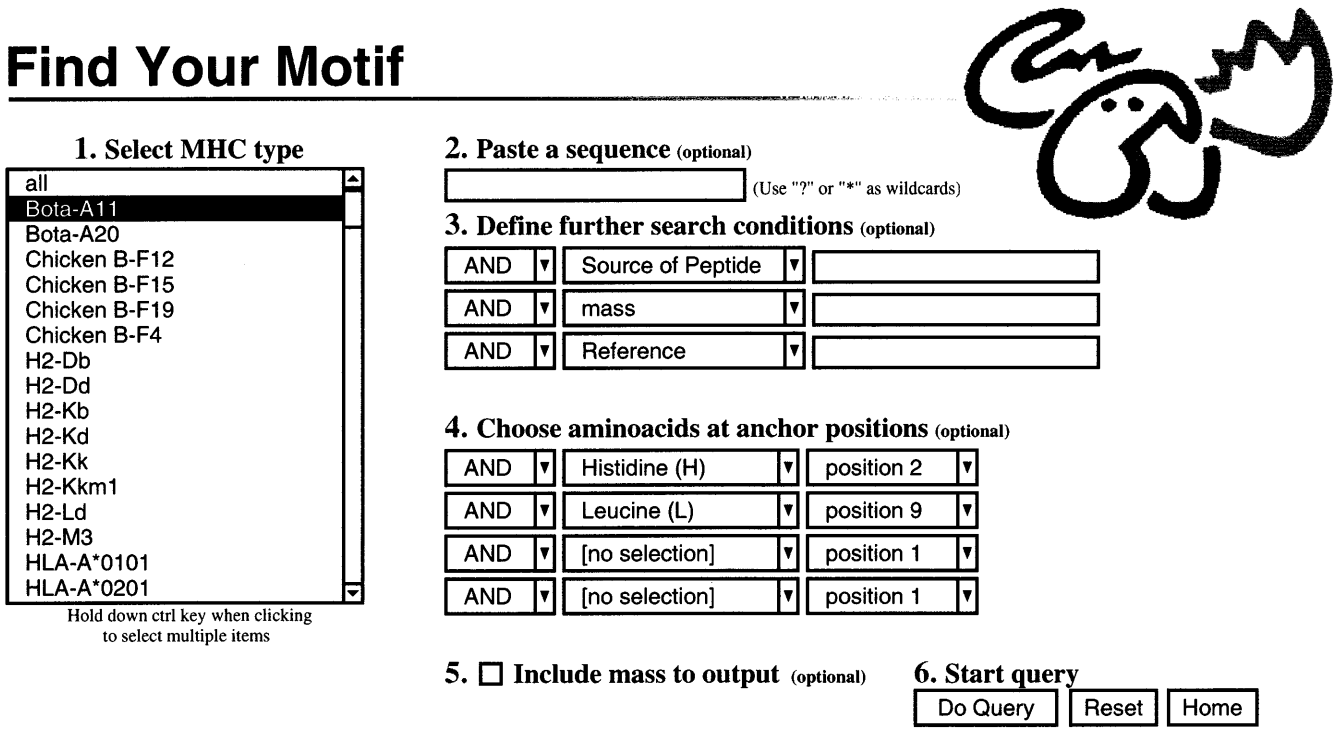

Fig. 1 'Find Your Motif' section of the SYFPEITHI database

tion on the specific role (anchor and auxiliary anchor amino acids) and position of each individual amino acid (aa). When data are browsed, the information is transformed to present a formatted version in which anchor aa are given in bold letters and auxiliary anchors are underlined. The main table is linked in a many-to-one relationship to the list of sources, and in a many-to-many relationship with the table of references. Each record in the table of source proteins refers to the specific EMBL ID, whereas each entry in the table of references is linked to the accession number (AN) of the reference in the NLM-PubMed database.

Table 1 Motif pattern for the prediction of HLA-B*1510 ligands

AA	1	**2**	3	4	5	6	7	8	**9**
A	0	0	1	0	0	0	0	1	0
C	0	0	0	0	0	0	0	0	0
D	0	0	0	1	0	0	0	0	0
E	1	0	1	1	0	0	0	1	0
F	0	0	0	0	0	0	0	0	**6**
G	1	0	0	1	1	0	0	0	0
H	0	**10**	0	0	0	0	0	0	0
I	2	0	0	0	0	1	0	0	0
K	0	0	0	1	0	1	0	0	0
L	0	0	0	0	0	0	0	0	**10**
M	0	0	0	0	0	1	0	0	**6**
N	0	0	0	0	1	0	0	0	0
P	0	0	0	2	1	1	1	0	0
Q	0	0	0	1	0	0	0	0	0
R	0	0	0	0	0	1	2	2	0
S	0	0	1	0	0	0	0	0	0
T	1	0	0	0	0	0	0	1	0
V	0	0	0	1	0	1	2	2	0
W	0	0	0	0	0	0	0	0	0
X	0	0	0	0	0	0	0	0	0
Y	1	0	0	0	0	0	0	0	0

Database retrieval can be performed on any HTML-browser supporting JavaScript. The main page of the database (http://www.uni-tuebingen.de/uni/kxi/) offers three sections: "Find Your Motif", "Epitope prediction" and "Information". After a preselection of one or multiple MHC-types, the "Find Your Motif" section (Fig. 1) allows the user to search for a complete or truncated sequence of up to nine aa, a given peptide source, or a reference. The search can be narrowed down even further by choosing a specific aa on a given position as anchor or auxiliary anchor. All search criteria may also be combined to obtain a complex analysis. When a search is performed, an SQL-query is generated and the results are presented on a dynamically composed HTML page. The page of results lists the MHC type, motifs, peptide sources, and references. From each peptide source and each reference, a hyperlink to the EMBL or PubMed database is generated, respectively.

The algorithm used for epitope prediction is written in Object-Pascal. In brief, a two-dimensional data array is built up, where the letters of the aa represent the row index and the pocket numbers represent the column index (Table 1). The scores in the array-cells of the matrix shown in Table 1 can be addressed directly by a pair of indices. Starting at the first aa, the sequence is then divided into octa-, nona- or decamers and for each oligomer the sum of the scores of the aa contained is calculated. The process is then repeated until the end of the sequence is reached. Amino acids that frequently occur in anchor positions are given the value 10, the value 8 is given to amino acids present in a significant number of ligands, and 6 for rarely occurring residues; amino acids of auxiliary anchor positions are given the value 6, less frequent residues of the same set have a coefficient of 4; preferred amino acids have coefficients of 1–4 according to the strength of signals in pool sequencing or the occurrence in individual sequences. Amino acids that are regarded as unfavorable for binding have a coefficient of –1 to –3. These values are taken into account in the algorithm.

Results and discussion

The main contents of the first SYFPEITHI release are shown in Table 3. Since the First Listing, a nearly five-

Table 2 Peptide motif and natural ligands of HLA-DRB1*0301

	Position	Source protein	Reference
	1 2 3 **4** 5 **6** 7 8 **9**		
Anchors			Malcherek et al. 1993; Geluk et al. 1992, 1994
	L D K Y		
	I R L		
	F E F		
	M Q		
	V N		
Examples for ligands			
	I S N Q **L** T L **D** S **N** T K **Y** F H K L N	Apolipoprotein B-100 (2877–2894)	Malcherek et al. 1993
	I S N Q **L** T L **D** S **N** T K **Y** F H K L	Apolipoprotein B-100 (2877–2893)	Malcherek et al. 1993
	I S N Q **L** T L **D** S **N** T K **Y** F H K	Apolipoprotein B-100 (2877–2892)	Malcherek et al. 1993
	V D T **F** L E **D** V **K** N L **Y** H S E A	α1-Antitrypsion (149–164)	Malcherek et al. 1993
	K P R A **I** V V **D** P V H G **F** M Y	LDL receptor (518–532)	Malcherek et al. 1993
	K Q T **I** S P **D** Y **R** N M I	IG2a (384–395)	Malcherek et al. 1993
	Y P D **F** I M **D** P **K** E K D K V	Unknown	Malcherek et al. 1993
	N I Q **L** I N **D** Q **E** V A R F D	Unknown	Malcherek et al. 1993
	L L S **F** V R **D** L **N** Q Y R A D I	Transferrin receptor (618–632)	Malcherek et al. 1993
L P K P P K P V S K **M** R M A T P L		Invariant chain (97–111)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
L P K P P K P V S K **M** R M A T P L L M Q A L P		Invariant chain (97–119)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
L P K P P K P V S K **M** R M A T P L L M Q A L P M		Invariant chain (97–120)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
P K P P K P V S K **M** R M A T P L		Invariant chain (98–113)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
P K P P K P V S K **M** R M A T P L L M Q A		Invariant chain (98–117)	Riberdy et al. 1992; Chicz et al. 1993: Sette et al. 1992
K P P K P V S K **M** R M A T P L L M Q		Invariant chain (99–116)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
K P P K P V S K **M** R M A T P L L M Q A L P M		Invariant chain (99–119)	Riberdy et al. 1992; Chicz et al. 1993; Sette et al. 1992
V D D T Q F **V** R F **D** S D A A S Q		HLA-A30 (52–67)	Chicz et al. 1993
A T K Y G N **M** T E **D** H V M H **L** L Q N A		Invariant chain (131–149)	Chicz et al. 1993
V F L L **L** L A **D** K V P E T S L S		ACh receptor (289–304)	Chicz et al. 1993
L N K **I** L L **D** E **Q** A Q W K		ICAM-2 (64–76)	Chicz et al. 1993
G P P K L D **I** R K E E **K** Q I M I D I F H		IFNγ receptor (128–147)	Chicz et al. 1993
G P P K L D **I** R K E E **K** Q I M I D I F H P		IFNγ receptor (128–148)	Chicz et al. 1993
G F K A **I** R P **D** K **K** S N P I I R T V		Cyt-B5 (155–172)	Chicz et al. 1993
Y A N **I** L L **D** R **R** V P Q T D M T F		Apolipoprotein B-100 (1207–1224)	Chicz et al. 1993
N L F **L** K S **D** G **R** I K **Y** T L N K N S L K		Apolipoprotein B-100 (1276–1295)	Chicz et al. 1993
I P D N L F **L** K S **D** G **R** I K **Y** T L N K N		Apolipoprotein B-100 (1273–1292)	Chicz et al. 1993
I P D N L F **L** K S **D** G **R** I K **Y** T L N K		Apolipoprotein B-100 (1273–1291)	Chicz et al. 1993
I P D N L F **L** K S **D** G **R** I K **Y** T L N		Apolipoprotein B-100 (1273–1290)	Malcherek et al. 1993; Chicz et al. 1993
I P D N L F **L** K S **D** G **R** I K **Y** T L		Apolipoprotein B-100 (1273–1289)	Chicz et al. 1993
N L F **L** K S **D** G **R** I K **Y** T L N K		Apolipoprotein B-100 (1276–1291)	Chicz et al. 1993
N L F **L** K S **D** G **R** I K **Y** T L N		Apolipoprotein B-100 (1276–1290)	Chicz et al. 1993
V T T **L** N S **D** L **K** Y N A L D L T N		Apolipoprotein B-100 (1294–1310)	Chicz et al. 1993
V G S **D** W **R** F L R G Y H Q Y A		HLA-A2 (103–117)	Chicz et al. 1993
T-cell epitopes			
G D V V A V V **D** I **K** E K G K D K W I E L K		Lol pol. P1 (171–190)	Geluk et al. 1994
K T **I** A Y **D** E **E** A R R		HSP65 (cattle) (3–13)	Hawes et al. 1995
M G R S I **K** V Q **L** Q		M. tuberculosis 30/31 kD protein (56–65)	Geluk et al. 1997
S D K N P **L** F L **D** E **Q** L I		M. tuberculosis HSP70 (257–269)	Geluk et al. 1997

fold increase in MHC motifs has been registered; the database SYFPEITHI now includes approximately 200 peptide motifs and 2000 peptide sequences. Each entry contains the peptide sequence, its MHC specificity, source protein, anchor positions, and publication references with links to the respective sequences in the EMBL databank and to the NLM-literature database PubMed. We refrained from including sequences other than those confirmed to ensure that only natural ligands and T-cell epitopes that are relevant for the respective MHC molecule are listed. A different database, MHCPEP (http://wehih.wehi.edu.au/mhcpep/), offered at the WEHI in Australia (Brusić et al. 1998) includes as many as 13 000 entries, in which submissions of preliminary data are included. MHCPEP also contains peptides that have been reported to bind to MHC in the absence of any functional data. Such peptides have been omitted from the SYFPEITHI database.

Table 4 depicts the MHC ligands of HLA-B*1510 with the anchoring amino acids printed in bold letters as an example of a database entry. If available, auxiliary anchors are underlined, and preferred residues are also given. The typical length of a class I ligand comprises 9 amino acids. Below the anchor positions a list of ligands and T-cell epitopes specific for the respective molecule follows. This includes the respective protein sources and references, which are directly linked to other databases available on-line. Every single MHC allele has its individual peptide specificity that is defined primarily by the position and specificity of the pockets that accommodate the side chains of the anchoring amino acids and in the second place by interactions of non-anchoring amino acid residues of the peptides.

Class II ligands consist of 12 to 25 amino acids, nine of which occupy the binding groove; between two and four are anchored in the pockets. As in the case of class

Table 3 MHC molecules currently included in the database

Class I

Human:

HLA-A1
HLA-A*0201
HLA-A*0202
HLA-A*0203
HLA-A*0204
HLA-A*0205
HLA-A*0206
HLA-A*0207
HLA-A*0209
HLA-A*0214
HLA-A3
HLA-A*0301
HLA-A*1101
HLA-A24
HLA-A*2402
HLA-A25
HLA-A26
HLA-A*2601
HLA-A*2602
HLA-A*2603
HLA-A29
HLA-A*2902
HLA-A*3001
HLA-A*3002
HLA-A*3003
HLA-A*3004
HLA-A*3101
HLA-A32
HLA-A*3301
HLA-A*3302
HLA-A*6601
HLA-A*6801
HLA-A*6802
HLA-A*6901
HLA-A*7401
HLA-B7
HLA-B*0702
HLA-B*0703
HLA-B*0705
HLA-B8
HLA-B*0801
HLA-B*0802
HLA-B13
HLA-B14
HLA-B15
HLA-B*1501 (B62)
HLA-B*1502
HLA-B*1503
HLA-B*1508
HLA-B*1509
HLA-B*1510
HLA-B*1513
HLA-B*1516
HLA-B*1517
HLA-B17
HLA-B18
HLA-B*1801
HLA-B22
HLA-B27
HLA-B*2702
HLA-B*2703
HLA-B*2704
HLA-B*2705
HLA-B*2706
HLA-B*2707
HLA-B*2709
HLA-B35
HLA-B*3501
HLA-B*3503
HLA-B37
HLA-B*3701
HLA-B*3801
HLA-B39
HLA-B*39011
HLA-B*3902
HLA-B*40011
HLA-B*40012 (B60)
HLA-B*4002
HLA-B*4006 (B61)
HLA-B42
HLA-B44
HLA-B*4402
HLA-B*4403
HLA-B*4405
HLA-B45
HLA-B*4501
HLA-B*4601
HLA-B*4801
HLA-B51
HLA-B*5101
HLA-B*5102
HLA-B*5103
HLA-B52
HLA-B*5201
HLA-B*5301
HLA-B*5401
HLA-B*5501
HLA-B*5502
HLA-B*5601
HLA-B57
HLA-B*5701
HLA-B*5702
HLA-B58
HLA-B*5801
HLA-B*5802
HLA-B*6701
HLA-B*7301
HLA-B*7801
HLA-Cw*0301
HLA-Cw*0304
HLA-Cw*0401
HLA-Cw*0601
HLA-Cw*0602
HLA-Cw*0702
HLA-Cw*1601
HLA-E
HLA-G

Macacca fascicularis:

Mafa-A2

Macacca mulatta:

Mamu-A*01
Mamu-A*08
Mamu-B*12

Pan paniscus:

Papa-Q*06

Pan troglodytes:

Patr-A*04
Patr-A*11
Patr-B*01
Patr-B*13
Patr-B*16

Saguinus oedipus:

Saoe-G*01
Saoe-G*06 or Saoe-G*04
Saoe-G*08
Saoe-G, unassigned

Mouse:

H2-D^{b}
H2-D^{d}
H2-K^{b}
H2-K^{d}
H2-K^{k}
H2-K^{km1}
H2-L^{d}
H2-M3
H2-M3f
Qa-1a
Qa-1b
Qa-2

Rat:

RT1.A^{a}
RT1.A^{a} (Tap2A-associated
RT1.A^{a} (Tap2B-associated)
RT1.A^{u}
RT1.A1^{c}
RT1.A^{l}

Cattle:

Bota-A11
Bota-A20

Chicken:

Chicken B-F12
Chicken B-F15
Chicken B-F19
Chicken B-F4

Pig:

D/d

Class II

Human:

HLA-DQ1
HLA-DQ6 or –DQ2
HLA-DQA1*0201/DQB1*0202
HLA-DQA1*0301/DQB1*0302 (DQ8)
HLA-DQA1*0301/DQB1*0303 (DQ9)
HLA-DQA1*0501/DQB1*0201(DQ2)
HLA-DQA1*0501/DQB1*0301(DQ7)
HLA-DQB1*0201
HLA-DQA1*0301/DQB1*0301 (DQ3.1)
HLA-DQA1*0301/DQB1*0401 (DQ4)
HLA-DQB1*0603
HLA-DR2
HLA-DR2 (DRB5*0101 or DRB1*1501)
HLA-DRB1*0101
HLA-DRB1*0102
HLA-DR17 or DRw52
HLA-DR3
HLA-DRB1*0301 (DR17)
HLA-DRB1*0401 (DR4Dw4)
HLA-DRB1*0402 (DR4Dw10)
HLA-DRB1*0403 (DR3Dw13)
HLA-DRB1*0404 (DR4Dw14)
HLA-DRB1*0405 (DR4Dw15)
HLA-DRB1*0406
HLA-DRB1*0407
HLA-DR7
HLA-DRB1*0701
HLA-DR8
HLA-DRB1*0801
HLA-DRB1*0802
HLA-DRB1*08032
HLA-DRB1*0901
HLA-DRB1*1001
HLA-DR11
HLA-DR11 or Dw52
HLA-DRB1*1101
HLA-DRB1*1102
HLA-DRB1*1104
HLA-DRB1*1201
HLA-DRB1*1301
HLA-DRB1*1302
HLA-DRB1*1401
HLA-DRB1*1405
HLA-DRB1*1501 (DR2b)
HLA-DRB1*1502
HLA-DRB1*1601
HLA-DRB1, unassigned
HLA-DRB3*0101
HLA-DRB3*0202 (DR52Dw25)
HLA-DRB3*0301 (DR52Dw26)
HLA-DRB4
HLA-DRB4*0101
HLA-DRB5*0101 (DR2a)
HLA-DRB5*0202
HLA-DRB, unassigned
HLA-DRw11
HLA-DRw52
HLA-DRw52 or HLA-DQ2

Macacca mulatta:

Mamu-DRB1*0406
Mamu-DRB*W201

Mouse:

H2-A^{b}, H2-A^{d}, H2-A^{g7}, H2-A^{k}, H2-A^{s}, H2-A^{u}, H2-E^{b}, H2-E^{d}

Rat:

RT1.B^{l}

Table 4 Peptide motif and natural ligands of HLA-B*1510 (Seeger and co-workers, in press)

	Position 1 **2** 3 4 5 6 7 8 **9**	**Source**	**Accession No. EMBL database**
Anchor residues	**H** **L**		
Preferred residues	I E P A V V V F Y A D G I R R M T S E P M P E G G N K T E K R A Q V		
Examples for ligands	G **H** D P R A Q G T **L**	HLA-DP α chain (220–229)	X00457
	D **H** C V A H K **L**	Cytochrome C reductase (66–73)	M36647
	I **H** E D S T N R R R **L**	Heat shock protein 90 b (440–450)	M16660
	E **H** A H N M R V **M**	Elongation factor 2 (489–497)	M19997
	G **H** L E N N P A **L**	60 S acidic rib. protein PQ (67–75)	M17885
	H **H** S G A K V V **L**	Chaperonin cont. TCP-1 η (282–290)	AF026292
	I **H** D P G R G A P **L**	60 S ribosomal protein L8 (49–58)	Z28407
	T **H** T Q P G V Q **L**	Septin 2 homologue (70–78)	D50918
	T **H** Y V A P R R **L**	Transcription activator SNF2L4 (899–907)	U29175
	Y **H** G H G V S A **F**	Human EST	T96718
	Y Q E K G V R V **L**	Actin-related protein Arp2 (402–410)	AF006082
	I **H** E P E P H I **L**	Cyclin-dep. kin. reg. subunit 1 (59–67)	X54941
	A **H** S T I M P R **L**	DNA repl. lic. factor MCM4 (694–702)	X74794
	E **H** A G V I S V **L**	HBV X interacting protein (40–48)	AF029890

I ligands, the nonanchoring amino acids play a secondary, but still significant role. A number of examples of ligands specific for HLA-DRB1*0301 are given in Table 2.

Information on the allelic specificity of the motifs, including preferred and unfavored residues beyond the anchor, enable the prediction of MHC ligands and T-cell epitopes. The second section of the database is dedicated to "Epitope prediction". The main objective of performing theoretical predictions is simply to save time. Instead of synthesizing and testing dozens or even hundreds of peptides, a preselection of a small set of peptides is made. The sequence of the protein or its gene, the restriction element, and its respective motif have to be available.

Motif-based predictions result in a list of peptides that have a high probability of being presented by MHC class I molecules. The prediction method used in the database is based on the principle of a building-up pattern (see Material and methods).The pattern used for the prediction of HLA-B*1510 nonamers is shown in Table 1. The values of all possible nonamers of a given sequence are added together and the optimal T-cell epitope is expected within the ten high-scoring peptides of each protein. Scores for all source proteins of nonamer peptides and predicted epitopes specific for HLA-B*1510 are given in Table 5. The correct and therefore naturally presented T-cell epitope from any source protein is assumed to appear among the top 2% of peptides in the high score list in more than 90% of predictions. However, theoretical approaches cannot always guarantee success. About 10% of the predictions are still unable to identify the respective epitope because not all details of the motifs are entirely known. Epitope predictions are only available at present for a small number of MHC alleles, since all predictions undergo several rounds of thorough cross-checking and a certain number of natural ligands and/or T-cell epitopes have to be available to ensure an accurate and reliable prediction (Table 6). The occurrence of unusual anchor amino acids can be compensated by refining the motif, but still a small percentage of epitopes do not act in accordance with any rules of the respective motif. In these cases, the theoretical approach will fail and experimental epitope mapping is unavoidable. SYFPEITHI is the only freely available MHC database in the internet apart from an HLA peptide binding prediction offered by the NIH (http://www-bimas.dcrt.nih.gov/cgi-bin/molbio/hla_bind/). The NIH prediction relies upon binding data (Parker et al. 1994) and estimates the half-time of dissociation of a given MHC-peptide complex. EpiMer (http://www.brown.edu/Research/TB-HIVLab/Epimatrix/epimerlink.html; Meister et al. 1995; Roberts et al. 1996) is a commercial site.

Other programs for the prediction of T-cell epitopes are distributed on disk (e.g., Motifs, D'Amaro et al. 1995; Davenport et al. 1995; Devereux et al. 1984; Fleckenstein et al. 1997; Hammer et al. 1994). Prediction patterns for class II are not yet included in our database due to the highly degenerate anchor positions in most MHC class II motifs. These degenerate anchor positions render the prediction of MHC class II ligands rather difficult. If all possible anchor residues of an MHC class II peptide motif are included within a search pattern, a very high number of possible natural MHC II ligands will be predicted. Only few of the class II peptide motifs encompass anchor residues specific

Table 5 Motif prediction of HLA-B*1510 self peptides and epitopes

Actin-related protein 2 (ARP2; human)

AA pos.	Score	Sequence
378	19	YQEKGVRVL
227	15	IEQEQKLAL
29	14	EHIFPALVG
60	14	EASELRSML
164	14	THICPVYEG

Septin 2 homologue (SEP2; human)

AA pos.	Score	Sequence
70	25	THTQPGVQL
156	22	GHSLKSLDL
371	22	LHQDEKKKL
129	17	EELKIRRVL
84	15	DLQESNVRL

60 S acidic ribosomal protein RQ (RLA0; human)

AA pos.	Score	Sequence
67	25	GHLENNPAL
80	19	PHIRGNVGF
241	18	IINGYKRVL
46	15	SLRGKAVVL
196	15	GSIYNPEVL

Elongation factor 2 (EF2; human)

AA pos.	Score	Sequence
489	24	EHAHNMRVM
356	18	IHLPSPVTA
147	17	IAERIKPVL
491	17	AHNMRVMKF
843	16	GLKEGIPAL

Transcription activator SNF2L4 (SN24; human)

AA pos.	Score	Sequence
899	27	THYVAPRRL
1009	24	RHMQAKGVL
620	23	IHVESGKIL
889	23	HHCKLTQVL
968	23	LHKVLRPFL
252	18	PHGMGGPNM
522	18	IEKERMRRL

Cyclin-dependent kinase regulatory subunit (CKS1; human)

AA pos.	Score	Sequence
59	26	IHEPEPHIL
16	14	EFEYRHVML
64	14	PHILLFRRP
29	13	AKLVPKTHL
38	13	MSESEWRNL

DNA replication license factor (MCM4; human)

AA pos.	Score	Sequence
694	25	AHSTIMPRL
331	21	CHTTHSMAL
255	20	EHQIQVRPF
440	16	LFSEKRVEL
741	16	AEAHAKVRL

enough to make motif-based predictions worthwhile, such as HLA-DRB1*0301, for example. In addition, anchor residues are often not used by class II ligands and epitopes. Inclusion of possibilities to predict class II ligands is in preparation.

Table 5 Continued

HIV1 GAG

AA pos.	Score	Sequence
192	22	GHQAAMQML
69	15	TGSEELRSL
143	15	HQAISPRTL
354	15	GPGHKARVL
42	14	RFAVNPGLL

Influenza A PR8/34 nucleoprotein

AA pos.	Score	Sequence
333	21	CHSAAFEDL
323	18	AHKSQLVWM
100	15	VNGKWMREL
336	15	AAFEDLRVL
48	14	KLSDYEGRL

Table 6 MHC motifs included in the database for epitope predicitions

Human	Mouse
HLA-A*0201	H2-D^{b}
HLA-A*0202	H2-K^{b}
HLA-A*0203	H2-K^{d}
HLA-A1	H2-K^{k}
HLA-A26	H2-L^{d}
HLA-B*0702	
HLA-B*1510	
HLA-B*2705	
HLA-B8	

Availability

SYFPEITHI is freely available online at the Interfakultäres Institut für Zellbiologie, Abteilung Immunologie, Tübingen (WWW site address: http://www.uni-tuebingen.de/uni/kxi).

Acknowledgments We acknowledge financial support of the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 510, Projekt E2) and the European Union and Biotech CT95-0263; Biomed CT95-1627. We also thank Vladimir Brusić for his valuable comments.

References

Brusić V, Rudy G, Kyne AP, Harrison LC (1998) MHCPEP, a database of MHC-binding peptides: update 1997. Nucleic Acids Res 26:368–371

Chicz RM, Urban RG, Gorga JC, Vignali DA, Lane WS, Strominger JL (1993) Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. J Exp Med 178:27–47

D'Amaro J, Houbiers JG, Drijfhout JW, Brandt RM, Schipper R, Bavinck JN, Melief CJ, Kast WM (1995) A computer program for predicting possible cytotoxic T lymphocyte epitopes based on HLA class I peptide-binding motifs. Hum Immunol 43:13–18

Davenport MP, Ho Shon IAP, Hill AVS (1995) An empirical method for the prediction of T-cell epitopes. Immunogenetics 42:392–397

Devereux J, Haeberli P, Smithies OA (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Falk K, Rötzschke O, Stevanović S, Jung G, Rammensee HG (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature 351:290–296

Fleckenstein B, Jung G, Wiesmüller KH (1997) Prediction and Design of new MHC class II-ligands based on the Activity Pattern of a Synthetic Undecapeptide Library. In: Brown F, Burton D, Doherty P, Mekalanos J, Norrby E (eds) Vaccines 97: Molecular approaches to the control of infectious diseases, Cold Spring Harbor Laboratory Press, pp 65–70

Geluk A, van Meijgaarden KE, Southwood S, Oseroff C, Drijfhout JW, de Vries RR, Ottenhoff TH, Sette A (1994) HLA-DR3 molecules can bind peptides carrying two alternative specific submotifs. J Immunol 152:5742–5748

Geluk A, van Meijgaarden KE, de Vries RR, Sette A, Ottenhoff TH (1997) A DR17-restricted T cell epitope from a secreted Mycobacterium tuberculosis antigen only binds to DR17 molecules at neutral pH. Eur J Immunol 27:842–847

Geluk A, Van Meijgaarden KE, Janson AA, Drijfhout JW, Meloen RH, De Vries RR, Ottenhoff TH (1992) Functional analysis of DR17(DR3)-restricted mycobacterial T cell epitopes reveals DR17-binding motif and enables the design of allele-specific competitor peptides. J Immunol 149:2864–2871

Hammer J, Bono E, Gallazzi F, Belunis C, Nagy Z, Sinigaglia F(1994) Precise prediction of major histocompatibility complex class II-peptide interaction based on peptide side chain scanning. J Exp Med 180:2353–2358

Hawes GE, Struyk L, Godthelp BC, van den Elsen PJ (1995) Limited restriction in the TCR-alpha beta V region usage of antigen-specific clones. Recognition of myelin basic protein (amino acids 84–102) and Mycobacterium bovis 65-kDa heat shock protein (amino acids 3–13) by T cell clones established from peripheral blood mononuclear cells of monozygotic twins and HLA-dientical individuals. J Immunol 154:555–566

Malcherek G, Falk K, Rötzschke O, Rammensee HG, Stevanovic S, Gnau V, Jung G, Melms A (1993) Natural peptide ligand motifs of two HLA molecules associated with myasthenia gravis. Int Immunol 5:1229–37

Meister GE, Roberts CGP, Berzofsky JA, De Groot AS (1995) Two novel T cell epitope prediction algorithms based on MHC-binding motifs; comparison of predicted and published epitopes from Mycobacterium tuberculosis and HIV protein sequences. Vaccine 13:581–91

Parker KC, Bednarek MA, Coligan JE (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J Immunol 152:163–175

Rammensee HG, Bachmann J, Stevanović S (1997) MHC ligands and peptide motifs. Landes Biosciences, Texas

Rammensee HG, Friede T, Stevanović S (1995) MHC ligands and peptide motifs: first listing. Immunogenetics 41:178–228

Riberdy JM, Newcomb JR, Surman MJ, Barbosa JA, Cresswell P (1992) HLA-DR molecules from an antigen-processing mutant cell line are associated with invariant chain peptides. Nature 360:474–477

Roberts CG, Meister GE, Jesdale BM, Lieberman J, Berzofsky JA, De Groot AS (1996) Prediction of HIV peptide epitopes by a novel algorithm. AIDS Res Hum Retroviruses 12:593–610

Seeger FH, Schirle M, Keilholz W, Rammensee HG, Stevanović (in press) Peptide motif of HLA-B*1510. Immunogenetics 49

Sette A et al. (1992) Invariant chain peptides in most HLA-DR molecules of an antigen-processing mutant. Science 258:1801–1804