242–244 *Nucleic Acids Research, 1996, Vol. 24, No. 1* *© 1996 Oxford University Press*

MHCPEP—a database of MHC-binding peptides: update 1995

Vladimir Brusic*, George Rudy, Anthony P. Kyne and Leonard C. Harrison

The Walter and Eliza Hall Institute, Parkville, Victoria 3050, Australia

Received September 8, 1995; Revised and Accepted October 2, 1995

ABSTRACT

MHCPEP is a curated database comprising over 6000 peptide sequences known to bind MHC molecules. Entries are compiled from published reports as well as from direct submissions of experimental data. Each entry contains peptide sequence, MHC specificity and when available, experimental method, observed activity, binding affinity, source protein, anchor positions, as well as publication references. The present format of the database allows text string matching searches but can easily be converted for use in conjunction with sequence analysis packages. The database can be accessed via Internet using Gopher, FTP or WWW.

INTRODUCTION

MHCPEP is a database of peptides known to bind MHC class I or II molecules. MHC binding is a prerequisite to T-cell recognition of peptide–MHC complexes, although not all binding peptides function as T-cell epitopes. Allele-specific motifs for peptides binding MHC class I are now well documented (1). Although motifs for peptides binding MHC class II have been reported (1–3), they are generally less well defined. The comprehensive compilation of binding peptides in MHCPEP enables the analysis of sequence properties governing binding to MHC molecules and the identification of T-cell epitopes. It should also facilitate research on antigen processing and transport, the mechanisms of T-cell receptor activation and the development of specific approaches to immunotherapy. Nearly 2500 new entries have been added since the last report (4). A number of duplicates from the original MHCPEP have been eliminated. Several entries have been retracted by contributing authors and subsequently removed. Formats of the entry ID and of the 'REFERENCES' and 'SUMMARY' fields have been changed.

DESCRIPTION

Version 1.1 of MHCPEP has 6413 entries (as of August 1995) compiled from published sources or directly submitted experimental data. A few peptides binding non-classical MHC molecules (e.g. mouse Qa-2a) are included. Each report of a peptide sequence is assigned to a separate entry identified by a unique entry ID. A given sequence may appear in several different entries, reflecting separately generated and reported data regarding that sequence. Entries consist of 12 fields composed of the field name followed by a colon (:) delimiter and the field value. Field values may be textual, numeric or empty. Fields are written one to a row and delimited by a hash (#). Entries are delimited by ellipses (...). Representative entries are shown in Figure 1. A description of each entry field follows:

Entry ID:

A unique identifier exists for each entry. The format is: [>organism][class][xxxx]. The 'organism' is designated by a three letter code (e.g. HUM for human, MUS for mouse); 'class' is a single digit number; the right side is an unique four digit hexadecimal number.

MHC molecule:

The designation of the MHC molecule, according to the nomenclature of Klein *et al.* (5), is followed by the specific allele (where known) within brackets. Human alleles are designated according to the DNA sequence nomenclature (6). This field also shows MHC class and host organism. The format is: [MHC molecule], [MHC class], [(host)].

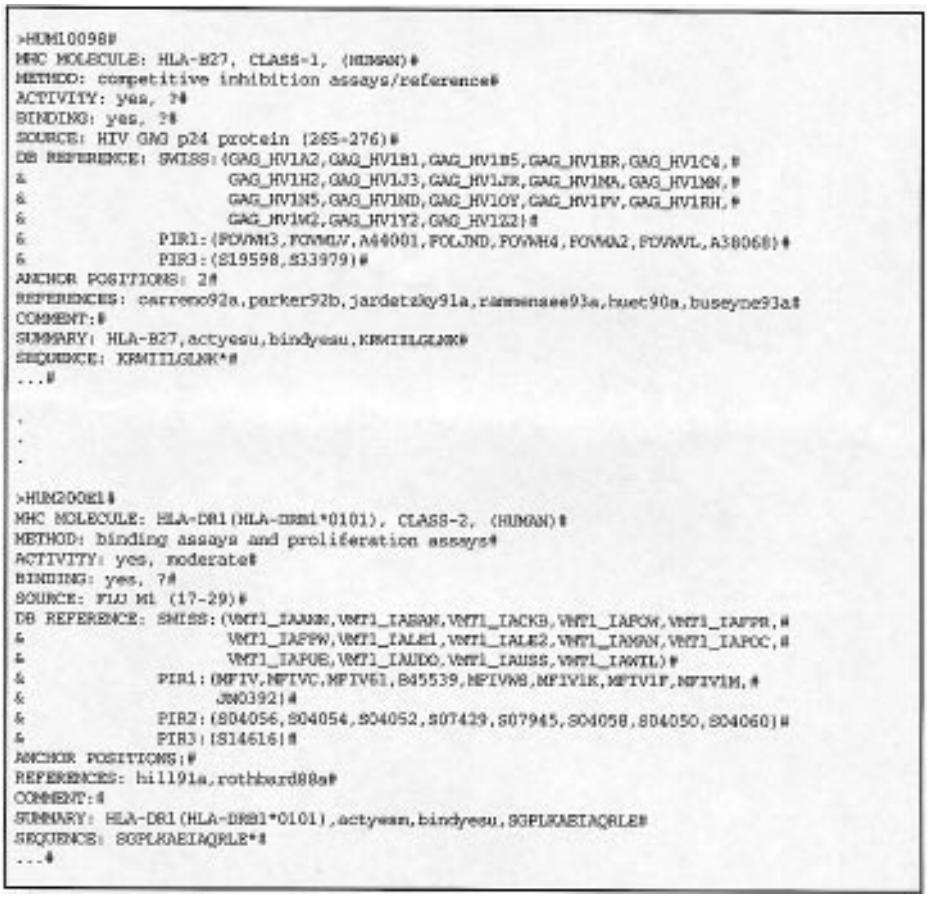
Method:

Peptide binding to MHC molecules is determined indirectly by T-cell based assays or directly by biochemical methods. T-cell recognition of MHC class I-bound peptides is usually detected by cytotoxicity assay, and of MHC class II-bound peptides by proliferation assay. Biochemical methods include direct binding, competitive inhibition assays, or elution followed by sequencing.

Activity:

The 'activity' of a peptide is a semi-quantitative measure of its immunogenic 'potency'. For a MHC class I-bound peptide, 'activity' is a measure of the extent of lysis by cytotoxic T-cells of target cells displaying the MHC class I-peptide complexes. A peptide is considered immunogenic if it mediates killing of at least 15% of the cells that display it. The 'activity' is expressed as the PD_{50}, (PD stands for peptide dose), the concentration of peptide giving 50% maximal specific lysis, and is given a descriptive value—none, little, moderate, high, immunogenic-not-quantified or unknown. A PD_{50} >10 μM is considered

* To whom correspondence should be addressed



```
>HUM10098#
MHC MOLECULE: HLA-B27, CLASS-1, (HUMAN)#
METHOD: competitive inhibition assays/reference#
ACTIVITY: yes, ?#
BINDING: yes, ?#
SOURCE: HIV GAG p24 protein (265-276)#
DB REFERENCE: SWISS:(GAG_HV1A2,GAG_HV1B1,GAG_HV1B5,GAG_HV1BR,GAG_HV1C4,#
&                   GAG_HV1H2,GAG_HV1J3,GAG_HV1JR,GAG_HV1MA,GAG_HV1MN,#
&                   GAG_HV1N5,GAG_HV1ND,GAG_HV1OY,GAG_HV1PV,GAG_HV1RH,#
&                   GAG_HV1W2,GAG_HV1Y2,GAG_HV122)#
&            PIR1:(POVWH3,POVWLV,A46001,POLJND,POVWH4,POVWA2,POVWVL,A38068)#
&            PIR3:(S19598,S33979)#
ANCHOR POSITIONS: 2#
REFERENCES: carreno92a,parker92b,jardetzky91a,rammensee93a,huet90a,buseyne93a#
COMMENT:#
SUMMARY: HLA-B27,actyeau,bindyeau,KRMIILGLNK#
SEQUENCE: KRMIILGLNK*#
...#

.
.
.

>HUM20061#
MHC MOLECULE: HLA-DR1(HLA-DRB1*0101), CLASS-2, (HUMAN)#
METHOD: binding assays and proliferation assays#
ACTIVITY: yes, moderate#
BINDING: yes, ?#
SOURCE: FLU M1 (17-29)#
DB REFERENCE: SWISS:(VM71_IAANN,VM71_IABAN,VM71_IACKB,VM71_IAPOM,VM71_IAPPR,#
&                   VM71_IAPPW,VM71_IALE1,VM71_IALE2,VM71_IAMAN,VM71_IAPOC,#
&                   VM71_IAPUE,VM71_IAUDO,VM71_IAUSS,VM71_IAWIL)#
&            PIR1:(MFIV,MFIVC,MFIV61,B45539,MFIVW8,MFIV1K,MFIV1F,MFIV1M,#
&                  JN0392)#
&            PIR2:(S04056,S04054,S04052,S07429,S07945,S04058,S04050,S04060)#
&            PIR3:(S14616)#
ANCHOR POSITIONS:#
REFERENCES: hill91a,rothbard88a#
COMMENT:#
SUMMARY: HLA-DR1(HLA-DRB1*0101),actyeau,bindyeau,SGPLKAEIAQRLE#
SEQUENCE: SGPLKAEIAQRLE*#
...#
```

Figure 1. Representative MHCPEP entries.

non-immunogenic and assigned ‘none’ as the value of the field ‘ACTIVITY’. For a MHC class II-bound peptide, ‘activity’ is a measure of the extent of T-cell proliferation induced by cells displaying the MHC class II-peptide complexes. Again ‘activity’ is expressed as PD_{50}, now defined as the concentration of peptide giving 50% maximal proliferation. The range of values of the field ‘ACTIVITY’ is given in Table 1.

Table 1. Range of values assigned to the field ‘ACTIVITY’

PD_{50}	Value
> 10 μM	none
10 μM–100 nM	yes, little
100 nM–1nM	yes, moderate
< 1nM	yes, high
immunogenic but unknown	yes, ?
immunogenicity unknown	?

Binding:

It is assumed that all ‘active’ peptides also bind; if no measure of binding is reported for an ‘active’ peptide, the value ‘yes, ?’ is assigned to the field ‘BINDING’. As several different methods exist for determining binding affinity, only a descriptive value is assigned to the data; the user should consult the original source for more specific details. The same scale as for ‘activity’ is used (none, little, moderate, high, unknown).

Source:

MHC peptides are fragments of larger proteins. This field indicates the parent protein with the start and end positions of the fragment. Synthetic peptides (e.g. those generated by mutations of a naturally occurring sequence) are designated by the word ‘homologue’.

DB reference:

This field specifies the name of the source protein(s) as it appears in the major protein databases: SWISS-PROT and PIR (PIR1, PIR2 and PIR3). These databases have been searched for sequences that match the MHC peptide entry sequence. This field may spread over several lines of text. The continuation of the field is designated by an ampersand (&) as the first character in the line.

Anchor positions:

Presumed anchor residues are numbered relative to the N-terminus of the peptide sequence. The main criterion for determining the values for this field is conformance to proposed binding motifs. A list of binding motifs compiled from the literature is provided with the database.

References:

A separate list of references to the published sources of entry sequences is supplied with the database. The value of the 'REFERENCES' field is a set of reference words, each of which consists of the first author's surname, year of publication or submission (two digit number), and an identifier (single letter). Each reference word uniquely designates a single reference. The format of the 'REFERENCES' field is: [author][year][x].

Comment:

This field is reserved for any relevant comments or observations.

Summary:

The summary field is a one-line description of the main fields of an entry (MHC molecule, activity, binding and peptide sequence), which is useful for rapid indexing of the database.

Sequence:

The sequence of the peptide is the one actually reported, not the minimum or optimum sequence. Therefore, a given T-cell epitope may be found within several entries representing different sequences which overlap or include it. The value for this field has an asterisk (*) following the C-terminal residue. The letter 'X' is used to represent ambiguity or an unknown residue. In some instances it is not possible to distinguish certain amino acids, e.g. tandem mass spectrometry does not distinguish leucine (L) and isoleucine (I). In such cases, separate entries are created and the ambiguity is noted in the 'COMMENT' field.

ACCURACY AND COMPLETENESS OF THE DATA

MHCPEP is largely compiled from published reports. However, numerous potential sources of error exist. Double-checking, comparison with original papers, comparison with other databases and multiple entry of the same sequence from different sources have been used to minimise errors. Some entries describing the same peptide may have different values in fields other than 'SEQUENCE' when derived from independent sources. Differences between T-cell clones used in experiments have not been considered; in cases where an MHC-bound peptide is recognised by any of the clones, it is entered into the database. Observations regarding clones which do or do not recognise the peptide are included in the comment field. The database has a degree of redundancy. This is unavoidable in order to include both peptides that are well-represented and those less well-represented in the literature. It also reflects the variety of ways of detecting MHC-bound peptides.

DATABASE ACCESS

MHCPEP is accessible via Internet using Gopher, FTP or WWW (described in ref. 7) to the following respective WEHI addresses:

gopher.wehi.edu.au
ftp.wehi.edu.au/pub/biology/mhcpep
http://www.wehi.edu.au

Text string searches are available using Sequence Retrieval System - SRS (8), accessible through the WEHI WWW home page. MHCPEP has been converted for local use with sequence analysis packages (e.g. GCG - ref. 9) which allows more sophisticated sequence analysis. A public access sequence analysis server is currently under development.

Authors who wish to cite MHCPEP should quote this text as the reference.

For queries and comments regarding the MHCPEP database contact (preferably by electronic mail): Vladimir Brusic, The Walter and Eliza Hall Institute, of Medical Research, PO Royal Melbourne Hospital, Victoria 3050, Australia; Tel: +61 3 9345 2588; FAX: +61 3 9347 0852; e-mail: vladimir@wehi.edu.au.

ACKNOWLEDGEMENTS

We are grateful to all those individuals who kindly submitted their data, and to Andrew Deavin for submission of the HLA class II T cell epitope catalogue. We would like to acknowledge the help of Ross Coppel, David Cram, Margo Honeyman, Andrew Lew and Nic Nicola during the course of designing and building the database. We appreciate the valuable help of Jo Marshall and Helen Ried in data collection.

REFERENCES

1 Rammensee, H.G., Friede, T. and Stevanovic S. (1995) *Immunogenetics*, **41**, 178–228.
2 Hammer, J., Valsasnini, P., Tolba, K., Bolin, D., Higelin, J., Takacs, B. and Sinigaglia, F. (1993) *Cell*, **74**, 197–203.
3 Chicz, R.M., Urban, R.G., Gorga, J.C., Vignali, D.A., Lane, W.S.and Strominger, J.L. (1993) *J.Exp.Med.*, **178**, 27–47.
4 Brusic, V., Rudy, G. and Harrison, L.C. (1994) *Nucleic Acids Res.*, **22**, 3663–3665.
5 Klein, J. Bontrop, R.E., Dawkins, R.L., Erlich, H.A., Gyllensten, U.B., Heise, E.R., Jones, P.P., Parham, P., Wakeland, E.K. and Watkins, D.I. (1990) *Immunogenetics*, **31**, 217–219.
6 Bodmer, J.G., Marsh, S.G., Albert, E.D., Bodmer, W.F., Bontrop R.E., Charron D., Dupont, B., Erlich, H.A., Mach, B., Mayr, W.R., Parham, P., Sasazuki, T., Schreuder, G.M.Th., Strominger, J.L., Svejgaard, A. and Terasaki, P.I. (1995) *Hum. Immunol.*, **43**, 149–164.
7 Krol, E. (1992) *The Whole Internet User's Guide & Catalog.* O'Reilly & Associates, Sebastopol, CA.
8 Etzold, T. and Argos, P. (1993) *CABIOS*, **9**, 49–57.
9 Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group, Madison, WI.