MHCBN: a comprehensive database of MHC binding and non-binding peptides

Manoj Bhasin, Harpreet Singh and G. P. S. Raghava*

Institute of Microbial Technology, Chandigarh, India

ABSTRACT

Summary: MHCBN is a comprehensive database of Major Histocompatibility Complex (MHC) binding and non-binding peptides compiled from published literature and existing databases. The latest version of the database has 19 777 entries including 17 129 MHC binders and 2648 MHC non-binders for more than 400 MHC molecules. The database has sequence and structure data of (a) source proteins of peptides and (b) MHC molecules. MHCBN has a number of web tools that include: (i) mapping of peptide on query sequence; (ii) search on any field; (iii) creation of data sets; and (iv) online data submission. The database also provides hypertext links to major databases like SWISS-PROT, PDB, IMGT/HLA-DB, GenBank and PUBMED.

Availability: MHCBN is available at http://www.imtech.res.in/raghava/mhcbn/. It's SRS version is available from http://srs.ebi.ac.uk/.

Contact: raghava@imtech.res.in

Supplementary information: http://www.imtech.res.in/raghava/mhcbn/supl/

INTRODUCTION

In the past, a number of databases have been created to provide information about MHC binding peptides and T cell epitopes. The databases like SYFPEITHI, JenPep, FIMM and HIV database are modest in size and provide precise information (Rammensee *et al.*, 1999; Blythe *et al.*, 2002; Schonbach *et al.*, 2002; Korber *et al.*, 1998). The MHCPEP is a widely used comprehensive database that contains information about 13 400 MHC binding peptides (Brusic *et al.*, 1998) but has got it's share of limitations. The main limitations of MHCPEP are (a) it has not been updated since 1998 and (b) it does not have tools for data extraction/analysis.

In order to overcome the limitations of MHCPEP and to provide the information about a large number of peptides, we have developed a comprehensive database, MHCBN, of MHC binders and non-binders. An overview of the database is represented in Figure 1. MHCBN

---

*To whom correspondence should be addressed.

contains detailed information about a peptide viz.: (i) amino acid sequence of peptide; (ii) associated MHC allele; (iii) sequence and structure of source protein; (iv) binding affinity in terms of IC_{50} value; (v) T-cell activity; (vi) sequence and structure of associate MHC molecules; (vii) experimental method; and (viii) source of information. The detailed description of each field of an entry/record which consists of information about a peptide is available at http://www.imtech.res.in/mhcbn/supl/. The major source of data for MHCBN is MHCPEP and the published literature. We have also included information from other immunological databases like FIMM, SYFPEITHI and HIV.

In this database, an attempt has been made to provide the information about reported MHC non-binders. These are the peptides which do not show binding affinity or show binding affinity below threshold in an experiment for a specific MHC allele. These non-binders are questionable because it is possible that in other experimental conditions, they may show binding affinity. However, the chances are much less that they show binding affinity. The non-binders are as important as binders for developing a method for predicting MHC binding peptide or to study interactions between the MHC molecule and the peptide. Thus, authors believe that it is worth maintaining the records for non-binders along with binders in the database.

The database also maintains 400 sequences and 106 structures of associated MHC molecules with hyperlinks to SWISS-PROT, GenBank, IMGT/HLA-DB, PDB etc. (Bairoch and Apweiler, 2000; Benson *et al.*, 2002; Robinson *et al.*, 2001). It also maintains 3296 sequences and 791 structures of source antigens/proteins with links to related databanks. The majority of database entries contain human and mouse MHC peptides. There are also a small number of entries containing MHC peptides of rat, chimpanzee, monkey, *Saguinus oedipus* and equine (http://www.imtech.res.in/raghava/mhcbn/supl/).

MHCBN provides a number of online web tools that allow users to retrieve and analyze the data. The following are the major tools: (i) general search allows user to search any field of database by using keywords with many options; (ii) peptide search allows the extraction
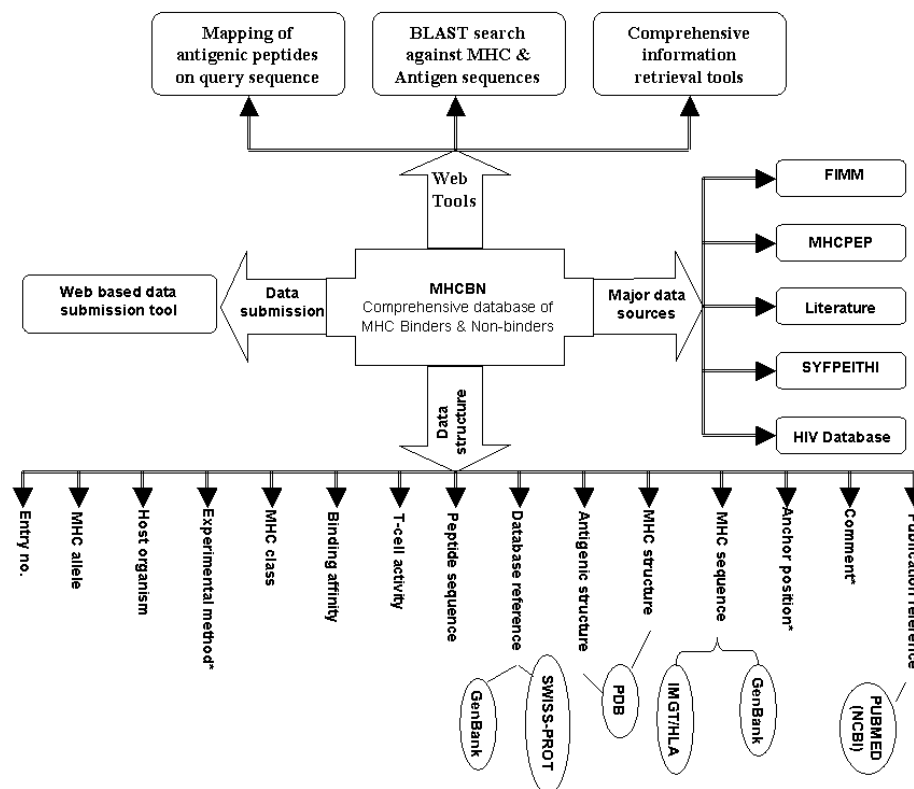
Fig. 1. Architecture of MHCBN database.

of peptides from database that contains the sequence identical to query peptide sequence or with a few allowed mismatches; (iii) tool for creation of data sets of MHC binders, non-binders, epitopes etc. for a given MHC allele; (iv) mapping of peptides facility allows mapping of experimentally determined MHC binders, non-binders and T-cell epitopes, on a query protein/antigen sequence; and (v) online data submission tool facility, allows to submit data to MHCBN via web. In addition, MHCBN has a facility to perform BLAST search of query protein sequence, against antigen and MHC sequences maintained at MHCBN (Altschul *et al.*, 1997).

ACKNOWLEDGEMENTS

REFERENCES

Altschul,S.F., Madden,T.L., Schäffer,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.

Bairoch,A. and Apweiler,R. (2000) The SWISS-PROT protein sequence database and its supplyment TrEMBL in 2000. *Nucleic Acids Res.*, **28**, 45–48.

Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J., Rapp,B.A. and wheeler,D.L. (2002) GenBank. *Nucleic Acids Res.*, **30**, 17–20.

Blythe,M.J., Doytchinova,I.A. and Flower,D.R. (2002) JenPep: a database of quantitative functional peptide data for immunology. *Bioinformatics*, **18**, 434–439.

Brusic,V., Rudy,G. and Harrison,L.C. (1998) MHCPEP, a database of MHC-binding peptides: update 1997. *Nucleic Acids Res.*, **26**, 368–371.

Korber,B.T.M., Moore,J.P., Brander,C., Walker,B.D., Haynes,B.F. and Koup,R. (1998) *HIV Molecular Immunology Compendium.* Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, NM.

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.

Robinson,J., Waller,M.J., Parham,P., Bodmer,J.G. and Marsh,S.G. (2001) IMGT/HLA database—a sequence database for the human major histocompatibility complex. *Nucleic Acids Res.*, **29**, 210–213.

Schonbach,C., Koh,J.L.Y., Flower,D.R., Wong,L. and Brusic,V. (2002) FIMM, a database of functional molecular immunology: update 2002. *Nucleic Acids Res.*, **30**, 226–229.