222–224 Nucleic Acids Research, 2000, Vol. 28, No. 1 

# FIMM, a database of functional molecular immunology

**Christian Schönbach, Judice L.Y. Koh, Xin Sheng, Limsoon Wong and Vladimir Brusic***

BIC/KRDL, Kent Ridge Digital Labs, 21 Heng Mui Keng Terrace, 119613 Singapore

Received September 20, 1999; Revised and Accepted October 27, 1999

## ABSTRACT

**FIMM database (http://sdmc.krdl.org.sg:8080/fimm ) contains data relevant to functional molecular immunology, focusing on cellular immunology. It contains fully referenced data on protein antigens, major histocompatibility complex (MHC) molecules, MHC-associated peptides and relevant disease associations. FIMM has a set of search tools for extraction of information and results are presented as lists or as reports.**

## INTRODUCTION

T cells of the immune system are involved in cell mediated immunity, regulation of immune responses and also provide help for production of antibodies (humoral immunity). T cells have T-cell receptors (TcR) which mediate recognition of antigenic structures and discrimination of self versus non-self. Peptides generated by processing of protein antigens bind MHC molecules, and are presented on the cell surface for recognition by TcR. MHC-associated peptides which induce T cell responses are termed T-cell epitopes.

The diversity of immune receptors arises by several mechanisms including phenotypic variation (MHC), allelic variation of genes (MHC, TcR), combining of chains in heterodimeric molecules (some MHC molecules, TcR), combinatorial joining of gene segments (TcR), or insertion of nucleotides at gene segment junctions (TcR). The MHC gene family is highly polymorphic—more than 700 allelic variants of human leukocyte antigens (HLA—human MHC) have been characterized (1). The diversity of immune system receptors allows an immune system to initiate and regulate appropriate responses. Hundreds of disease-specific antigens have been identified and reported. Thousands of peptides have been reported to bind various MHC molecules or stimulate immune responses (2,3). Sets of peptides that are presented by different MHC molecules may overlap to various degrees, or may be exclusive. Associations between HLA genes and susceptibility (or protection) to diseases have been reported (4). This complexity created a need for a database that integrates data on functional aspects of molecular immunology.

FIMM contains fully referenced data on protein antigens, major histocompatibility complex (MHC) molecules, MHC-associated peptides and relevant disease associations. A set of search and querying tools allows users to perform specific queries and combine different views of data. Extracted information is in the form of reports or lists containing hyperlinks to other sources that provide more detailed or specialized information. The reports and lists are designed to facilitate data interpretation and help design related experiments. Data in FIMM originate from various sources including literature, public databases and HLA workshop reports. FIMM is designed to assist both basic and applied research in molecular immunology. FIMM (version 1.0) was established in 1999 and contains data on more than 400 protein antigens, 1200 peptides, 800 HLA sequences, 50 diseases, 20 disease associations and 2000 references.

## DESCRIPTION

The purpose of the FIMM is to provide: (i) a unique compilation of information relevant to molecular immunology, (ii) means for extraction of this information, including the analysis of query antigens, and (iii) access by hyperlinks to related information available elsewhere. The dimensional data model (5) of FIMM is given in Table 1. The current FIMM data model has five dimensions (or views): protein antigens, peptides, MHC, diseases and publication sources. FIMM can be queried for specific information within a particular view. A set of generic tools allows keyword searches and sequence comparison analysis. Online documentation provides help for use and the description of the database. In addition to internal links, FIMM provides a rich set of hyperlinks to relevant external sites (Fig. 1).

### Common fields

Each factual entry in four data dimensions (except ‘Sources’ which have PubMed identifiers or PMIDs) has a FIMM unique identifier (ID) and the date of entry (Date). Factual entries for ‘Protein antigens’, ‘MHC’ and ‘Diseases’ also have the name of the entry (Name) and the list of alternative names (Aliases). Amino acid sequence information (Sequence) is available for ‘Protein antigens’, ‘Peptides’ and ‘MHC’.

### Protein antigens

Proteins that are reported either as sources of T-cell epitopes or of naturally processed peptides, are factual entries for the dimension ‘protein antigens’. For each entry, the most relevant sequence is provided along with links to related entries from the SWISS-PROT (6) and GenBank (7) databases. Sequence features contain information extracted from the descriptions in the SWISS-PROT or GenBank database entries.

### Peptides

Peptides reported as naturally processed and presented by MHC molecules or those reported as T-cell epitopes are factual entries for the dimension ‘Peptides’. Besides common fields, each entry contains information on peptide/MHC association

*To whom correspondence should be addressed. Tel: +65 874 7920; Fax: +65 774 8056; Email: vladimir@krdl.org.sg

**Table 1.** Data dimensions in FIMM

| Dimensions | | | | |
|---|---|---|---|---|
| Protein antigens | Peptides | MHC | Diseases | Sources |
| ID | ID | ID | ID | PMID |
| Date | Date | Date | Date | Title |
| Name | Sequence | Name | Name | Authors |
| Aliases | MHC | Aliases | Aliases | Journal |
| Features | T-cell epitope | Sequence | Etiology | Links |
| Sequence | Naturally processed | Disease assoc. | Disease assoc. | -PubMed |
| Links | Peptide binding | Binding pockets | Links | |
| -Peptides | Links | Binding sites CD8 | -Protein antigens | |
| -Diseases | -Protein antigens | Binding sites TcR | -MHC | |
| -Sources | -MHC | Alignments | -Sources | |
| -SWISS-PROT | -Sources | Sources | -OMIM | |
| -GenBank | -MHCPEP | Links | -GeneCards | |
| | -SYFPAETHY | -Peptides | -Genatlas | |
| | -HIVdB | -Diseases | | |
| | -SWISS-PROT | -IMGT database | | |

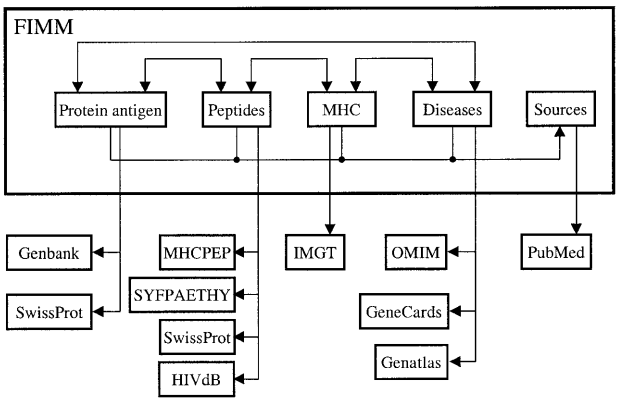


**Figure 1.** Data views in FIMM, internal links and links to the external sources.

and whether the peptide is naturally processed or a reported T-cell epitope. A peptide entry contains the information on peptide binding affinity when available, using the notation from the MHCPEP database (high, moderate or low binding affinity). Major sources of peptide information are the MHCPEP database (3; http://wehih.wehi.edu.au/mhcpep ), the SYFPEITHI database (H.G.Rammensee, *et al.*: SYFPEITHI: An Internet Database for MHC Ligands and Peptide Motifs, http://www.uni-tuebingen.de/uni/kxi/ ), HIV molecular immunology database (8; http://hiv-web.lanl.gov/immunology ) and published reports.

### MHC view

Factual entries of MHC contain the common fields (as described earlier in the text) and the publication sources. In addition the 'MHC' dimension contains information on MHC binding pocket composition (9), co-receptor (CD8) binding sites (10) and TcR binding sites (11). Binding pocket analysis allows the selection and inspection of binding pockets for a subset of MHC alleles for a specific phenotype. The major source of MHC sequence information for FIMM version 1.0 is the IMGT database (12; http://www.ebi.ac.uk/IMGT ).

### Diseases

A factual entry from the dimension 'Diseases' contains the following information: common fields, etiology (e.g., virus, cancer, etc.) and MHC disease association. External links include OMIM (13; http://www.ncbi.nlm.nih.gov/omim/ ), GeneCards (14; http://www.dkfz-heidelberg.de/GeneCards/ ) and Genatlas (15; http://web.citi2.fr/GENATLAS/ ) databases.

### Sources

The 'Sources' dimension contains publication references, namely the author name(s), title, journal and the links to PubMed (http://www.ncbi.nlm.nih.gov/PubMed/ ) records. Users can search references by keyword, author name(s) or title words.

### Search tools

FIMM integrates several tools for data searching. Searches using keywords (both full and partial) are available in each dimension. FIMM allows the creation of lists of related entries. The resulting lists contain summaries of factual entries (e.g., diseases, protein antigens, MHC, peptides and reference sources) generated according to various grouping criteria. Grouping criteria are usually defined by user-provided keywords. Alternatively, specific grouping criteria are also predefined in FIMM menus (e.g., list HLA alleles by loci, diseases by etiology, peptides by MHC molecules, etc.).

Users can compare their query sequence to FIMM entries using BLAST2 (16) searches. A query sequence can be compared with either protein antigens or MHC sequences in FIMM. Sequence alignments are performed using ClustalW (17) and the alignment coloring by MView (18). Peptides in FIMM can be searched by the motif search program 'pattern_find' which uses Perl regular expression techniques (see Acknowledgements). Binding pocket analysis can be performed for a query MHC sequence or with FIMM entries only. This search will provide the alignment of amino acids that form specific pockets in MHC molecules.

## DATA ANNOTATION

FIMM is compiled mainly from published reports and public databases. Cross-checking of data for both accuracy and redundancy is performed routinely. The inclusion of HLA sequences is based on publications by the HLA nomenclature committee (1). The criteria for the inclusion of a peptide into FIMM are that it has a complete sequence and has been previously published. Disease associations are compiled from selected publications, such as HLA workshop proceedings (19), or journal articles.

## DATABASE ACCESS

FIMM is available via WWW site: http://sdmc.krdl.org.sg:8080/fimm . FIMM has been designed to be robust and user-friendly. The user interface uses a set of simple Graphical User Interface forms. The data are stored in flat (non-structured), comprehensible and easy-to-access database files. Methods for searching the databases and displaying selected tables are built with a combination of Java, Perl, CGI and C programs. Development of the FIMM system was carried out in a UNIX environment.

Authors whose research has been assisted by using FIMM should cite this article as the reference.

## FUTURE WORK

Version (1.0) of FIMM contains information on human MHC (HLA) and human diseases. Future work will include relevant information from other organisms including laboratory and agricultural animals. Additional data dimensions and facilities planned for future FIMM developments include antigen processing, TcR interactions, cytokines and various prediction tools.

## ACKNOWLEDGEMENTS

We thank K. Hofmann of ISREC, Switzerland for making program pattern_find publicly available. We also thank Edward J. Altman of KRDL for useful discussions.

## REFERENCES

1. Bodmer,J.G., Marsh,S.G., Albert,E.D., Bodmer,W.F., Bontrop,R.E., Dupont,B., Erlich,H.A., Hansen,J.A., Mach,B., Mayr,W.R., Parham,P., Petersdorf,E.W., Sasazuki,T., Schreuder,G.M., Strominger,J.L., Svejgaard,A. and Terasaki,P.I. (1999) *Hum. Immunol.*, **60**, 361–395.
2. Rammensee,H.G., Friede,T. and Stevanovic,S. (1995) *Immunogenetics*, **41**, 178–228.
3. Brusic,V., Rudy,G. and Harrison,L.C. (1998) *Nucleic Acids Res.*, **26**, 368–371.
4. Thorsby,E. (1997) *Hum. Immunol.*, **53**, 1–11.
5. Ballard,C., Herreman,D., Schau,D., Bell,R., Kim,E. and Valencic,A. (1998) *Data Modeling Techniques for Data Warehousing*. IBM International Support Organization, San Jose, CA.
6. Bairoch,A. and Apweiler,R. (1999) *Nucleic Acids Res.*, **27**, 49–54. Updated article in this issue: *Nucleic Acids Res.* (2000), **28**, 45–48.
7. Benson,D.A., Boguski,M.S., Lipman,D.J., Ostell,J., Ouellette,B.F., Rapp,B.A. and Wheeler,D.L. (1999) *Nucleic Acids Res.*, **27**, 12–17. Updated article in this issue: *Nucleic Acids Res.* (2000), **28**, 15–18.
8. 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.
9. Chelvanayagam,G. (1997) *Hum. Immunol.*, **58**, 61–69.
10. Sullivan,J.A., Oettinger,H.F., Sachs,D.H. and Edge,A.S. (1997) *J. Immunol.*, **159**, 2318–2326.
11. Hashimoto,K., Okamura,K., Yamaguchi,H., Ototake,M., Nakanishi,T. and Kurosawa,Y. (1999) *Immunol. Rev.*, **167**, 81–100.
12. Lefranc,M.P., Giudicelli,V., Ginestoux,C., Bodmer,J., Muller,W., Bontrop,R., Lemaitre,M., Malik,A., Barbie,V. and Chaume,D. (1999) *Nucleic Acids Res.*, **27**, 209–212. Updated article in this issue: *Nucleic Acids Res.* (2000), **28**, 219–221.
13. McKusick,V.A. (1998) *Mendelian Inheritance in Man. Catalogs of Human Genes and Genetic Disorders*. Johns Hopkins University Press, Baltimore, MD.
14. Rebhan,M., Chalifa-Caspi,V., Prilusky,J. and Lancet,D. (1998) *Bioinformatics*, **14**, 656–664.
15. Frezal,J. (1998) *C. R. Acad. Sci. III*, **321**, 805–817.
16. Altschul,S.F. and Gish,W. (1996) *Methods Enzymol.*, **266**, 460–480.
17. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) *Nucleic Acids Res.*, **22**, 4673–4680.
18. Brown,N.P., Leroy,C. and Sander,C. (1998) *Bioinformatics*, **14**, 380–381.
19. Charron,D. (ed.) (1997) *Proceedings of the Twelfth International Histocompatibility Workshop and Conference*. EDK, Paris, France.