

An Update on the Functional Molecular Immunology (FIMM) Database

Christian Schönbach,^{1,2} Judice L.Y. Koh,¹ Darren R. Flower³ and Vladimir Brusic^{1,4}

1 Knowledge Discovery Department, Institute for Infocomm Research, Singapore

2 Immunoinformatics Team, Advanced Genome Information Technology Group, RIKEN Genomic Sciences Center, RIKEN Yokohama Institute, Yokohama, Japan

3 The Edward Jenner Institute for Vaccine Research, Compton, UK

4 Department of Microbiology, National University of Singapore, Singapore

Abstract

Data on the major histocompatibility complex, T-cell epitopes, B-cell epitopes, antigens and diseases are heterogeneous and scattered among different databases and the literature. Since it has become increasingly difficult to obtain an integrated view of functional immune response components, we have developed and updated over several years the Functional molecular IMMunology (FIMM) database (<http://research.i2r.a-star.edu.sg/fimm/>). FIMM contains integrated expert-curated data on protein antigens, and on human immunological receptors that recognise and bind them in healthy or disease states. Interfaces with multiple, intuitive query options and query reports provide immunologists with prioritised information that aids data interpretation, vaccine target discovery and immune disease research.

T cells of the immune system are involved in cell-mediated immunity, regulation of immune responses and regulation of antibody production by B cells (humoral immunity). T cells recognise antigenic structures (e.g. peptides) presented by class I and II molecules of the major histocompatibility complex (MHC) using T-cell receptors (TCR). The TCR can discriminate between antigenic self and foreign peptide-MHC structures. MHC-associated peptides that induce T-cell responses are termed T-cell epitopes. B cells proliferate, differentiate and raise antibody responses against T-cell-dependent antigens, if they are stimulated by certain populations of T cells. During differentiation, B cells produce immunoglobulins (Ig) antibodies that have unique structures to recognise and bind structural determinants on the surface of antigens. The antigenic determinants are called B-cell epitopes. Among several types of antibodies, IgG is the most common in humoral immunity, whereas IgE is associated with allergic immune responses.

The diversity of immune receptors (MHC, TCR and Ig) encompasses both phenotypic and genotypic variations. Variation of haplotype and gene level results in a variable number of genes in different haplotypes and extreme allelic polymorphism of individual genes, respectively. More than 1800 allele sequences of human leukocyte antigens (HLA; human MHC) have been characterised

and officially named,^[1] and compiled in the IMGT/HLA database.^[2]

There is the potential to create $>2 \times 10^{12}$ different antigen receptors per individual (10^{12} Ig^[3] and 10^{12} TCR^[4]), based on the combinatorial diversity of the Ig and TCR V-D-J (variable-diversity-joining region) genes, N-diversity at the V-D-J junctions of Ig-heavy (IGH) and TCR genes, somatic hypermutations of the rearranged Ig genes, and combination of chains (Ig heavy and light chains, TCR- α and - β chains, TCR- γ and - δ chains). Over 680 human Ig and TCR genes and 1200 alleles have been characterised and entered in IMGT/LIGM databases,^[5] and the corresponding rearranged sequences have been fully annotated in IMGT/LIGM-DB.^[6,7]

The diversity of immune system receptors allows the immune system to initiate and regulate appropriate responses. Thousands of peptides have been reported to bind various MHC molecules or stimulate T-cell responses and these are compiled in the databases MHCPEP,^[8] SYFPEITHI^[9] and JenPep,^[4] among others (see table I). Sets of peptides that are presented by different MHC molecules may overlap to various degrees or may be exclusive.

Associations between HLA genes and disease susceptibility or protection have been described by Thorsby,^[10] and hundreds of disease-specific antigens have been identified and reported in the literature. Only a fraction of the knowledge, mostly related to

Table 1. Overview of immunological databases

Database	Content	URL
Peptides and T-cell epitopes		
SYFPEITHI 1.0	>4000 curated T-cell epitopes, eluted MHC peptide ligand entries; MHC-binding motifs	http://www.syfpeithi.de/
HLA Ligand/Motif Online Database	Curated MHC ligands, T-cell epitope data and binding motifs	http://hlaligand.ouhsc.edu/
MHCPEP	>13 000 curated T-cell epitope and MHC-binding data	http://wehih.wehi.edu.au/mhcpep (now defunct; only FTP of data)
Peptides, T- and B-cell epitopes		
JenPep 2.0	>8000 curated quantitative and qualitative data on MHC ligands, T- and B-cell epitopes, TAP ligands and TCR-pMHC interactions	http://www.jenner.ac.uk/jenpep2/
AntiJen 2.0	>24 000 curated entries; extension of JenPep including immunological protein-protein interactions and peptide libraries	http://www.jenner.ac.uk/AntiJen/
Disease-specific epitopes and immunologically characterised tumour cell lines		
HIV Molecular Immunology	Curated MHC-binding motifs, ligands, B-cell (>500) and T-cell (>300) HIV epitopes	http://www.hiv.lanl.gov/content/immunology/index.html
Cancer Immunome Database	Serologically identified tumour antigens, antibody and tissue data	http://www2.licr.org/CancerImmunomeDB
FARRP (Food Allergy Research and Resource Program) Protein Allergen Database	>650 curated known and putative allergens compiled from literature	http://www.allergenonline.com/
Protall	>70 plant food allergens	http://www.ifr.bbsrc.ac.uk/protall/
SDAP (Structural Database of Allergenic Proteins)	Allergenic proteins with structural data (28) or models (1) and IgE epitopes (20)	http://fermi.utmb.edu/SDAP/
Allergen Nomenclature	Allergens and official nomenclature	http://www.allergen.org/
IPD-ESTDAB (European Searchable Tumour Line Database)	HLA-typed and immunologically characterised tumour cell lines	http://www.ebi.ac.uk/ipd/estdab/
MHC, TCR, Ig and related proteins of the immune system		
IMGT/LIGM-DB	Ig and TCR sequences from human and other vertebrate species; 82 800 sequences of 150 species	http://imgt.cines.fr/cgi-bin/IMGTlect.jv
IMGT/HLA	HLA sequences and nomenclature	http://www.ebi.ac.uk/imgt/
IMGT/PRIMER-DB	Oligonucleotides of the Ig, TCR and RPI; 1800 entries	http://imgt3d.igh.cnrs.fr/PrimerDB/Query_PrDB.pl?origin=view&Default=ok
IMGT/GENE-DB	Genome database for Ig and TCR genes (1375) and alleles (2201)	http://imgt.cines.fr/cgi-bin/GENEelect.jv?livret=0
IMGT/3Dstructure-DB	3-D structures for Ig, TCR, MHC and RPI with annotated sequences and IMGT Colliers de Perles (795 entries)	http://imgt3d.igh.cnrs.fr/
Kabat Database	Ig, TCR, MHC, complement, adhesion molecules and other immunologically relevant sequences	http://www.kabatdatabase.com/

Continued next page

Table I. Contd

Database	Content	URL
IPD-MHC	MHC sequences, nomenclature and phylogenetics of ape, monkey, cattle, dog and cat	http://www.ebi.ac.uk/ipd/mhc/
IPD-KIR	Killer-cell immunoglobulin-like receptor sequences, nomenclature and haplotypes	http://www.ebi.ac.uk/ipd/kir/
IPD-HPA	Human platelet antigen information including alloantigens	http://www.ebi.ac.uk/ipd/hpa/
Integrated data		
FIMM (Functional molecular IMMunology) 1.3	>85 000 entries on protein antigens, diseases, B- and T-cell epitopes, MHC, 3-D structures and models, references to literature	http://research.i2r.a-star.edu.sg/fimm/

3-D = 3-dimensional; **HLA** = human leukocyte antigen; **Ig** = immunoglobulin; **IMGT** = international ImMunoGeneTics information system®; **IPD** = Immuno Polymorphism Database; **MHC** = major histocompatibility complex; **pMHC** = peptide-MHC; **RPI** = related proteins of the immune system; **TAP** = transporter associated with antigen processing; **TCR** = T-cell receptor.

cancer and viral infections, has been integrated as context information into epitope databases.

With the increased incidence of allergic diseases such as asthma and atopic dermatitis, the number of allergenic determinants is also on the rise. However, existing allergen databases cover only a few, sometimes redundant, protein allergens,^[11] because the identification and classification of B-cell epitopes is a nontrivial task. Difficulties arise from (a) the spatial limitation of antibody accessibility to the allergen, (b) a small number of linear (formed of a contiguous sequence) epitopes compared with conformational (formed of non-contiguous segments) epitopes and (c) recognition of multiple different shapes by a single antibody.^[12]

As most of the existing immunology databases support only a small area of functional immunology, we developed the Functional molecular IMMunology (FIMM) database to integrate epitope and immune receptor sequences from the literature, public databases and HLA workshop reports^[13] with context information on immune diseases, antigens and HLA structures. Over the last 4 years, FIMM was expanded in data volume, content, complexity and functionality to facilitate data interpretation and planning of related experiments for understanding the initiation and regulation of immune response, vaccine design, development of immunotherapies and deimmunisation of proteins.

Data Model

A dimensional data model^[14] was used to organise the data in a simple way and to develop relevant query functions in discussion with immunologists. Details of the data structure of each dimension are available at <http://research.i2r.a-star.edu.sg/fimm/information.html#dimension>. The current FIMM data model (figure 1) has six dimensions (or views): protein antigens, peptides, MHC,

structures, diseases and publication references. FIMM views contains fully referenced data on protein antigens, MHC molecules, T-cell and B-cell (mostly IgE-binding) epitopes, naturally processed MHC-associated (self-) peptides and relevant annotated disease associations. Hyperlinks provide access to related information available in the external sources.

The entries of all FIMM dimensions are formatted as semi-structured flat files and interconnected through look-up indices. This data model permits search/display of epitope and/or natural ligand data from any dimension through a set of simple and advanced graphical user interfaces. Programs for searching the databases and displaying the content of selected entries are written with a combination of Java™, Perl, HTML, C shell scripts and C. We have migrated the updated FIMM database to a 2 × 296 MHz CPU/1.5GB memory UNIX® server. The new URL of the FIMM database is <http://research.i2r.a-star.edu.sg/fimm/>.

Functional Molecular Immunology (FIMM) Database Dimensions and Data

The growth of FIMM in both the number of data views (dimensions) and the amount of data is shown in table II. The new entries, since the previous release,^[15] include 61 new diseases (19 allergies, 42 cancers), 113 HLA-binding peptides, 266 IgE-binding B-cell epitopes of 37 allergens, 84 other antigens, 256 HLA sequence entries, 78 HLA class II 3-dimensional (3-D) homology models, 70 annotated HLA crystal structures derived from 20 HLA allele sequences, and 951 references. The current FIMM (version 1.3) contains immunological and disease-related data on 692 protein antigens, 1970 peptide entries, 1589 HLA entries, 627 HLA 3-D models and crystal structures, 3766 references and descriptions of 178 diseases.

Common Fields

Each factual entry in the four data dimensions of 'protein antigens', 'peptides', 'MHC' and 'diseases' has a FIMM unique identifier (ID) and the date of entry (Date). The 'references' use PubMed® identifiers or PMIDs, and the 'structures' use HLA allele names as identifiers. 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'.

Peptides

Peptides that are reported as naturally processed and presented by MHC molecules or as T-cell or B-cell epitopes are factual entries of the dimension 'peptides'. Each peptide entry contains the common fields, plus information on peptide/MHC allele associations, whether the peptide is naturally processed, and the reported sequence of its T-cell or B-cell epitope. MHC/peptide associations are allele based. The criteria for the inclusion of a peptide within FIMM are that it has (i) a complete sequence, (ii) been previously published and (iii) a valid MHC restriction on allele level. Peptides associated with obsolete alleles are deleted. For example, HIV-1 GAG epitope KYKLKHIVW was removed because it was reported with an erroneous HLA-A*2401 sequence.

B-cell epitope entries are currently restricted to allergens. The epitopes are annotated with the Ig-binding type (e.g. IgE, IgG) instead of the peptide/MHC association. Since the Ig epitope data

are derived from different experimental methods, expert curators simplified the data by classifying the epitopes into 'dominant', 'moderate' and 'weak'. Unclassifiable epitopes are labelled with 'yes', indicating allergenicity but lack of quantitative Ig-binding data.

MHC-binding peptide entries contain information on peptide binding affinity using, if available, the notation introduced in the MHCPEP database (high, moderate or low binding affinity). Peptides that represent T-cell epitopes are annotated as such. Major sources of peptide data are the MHCPEP database, SYFPEITHI database, HIV Molecular Immunology database^[16] and journal articles. Each peptide entry is hyperlinked to related entries in the databases SYFPEITHI, JenPep and HLA Ligand/Motif Online Database.^[17]

Protein Antigens

Proteins that are reported as sources of T-cell or B-cell epitopes or naturally processed HLA-binding peptides are factual entries for the dimension 'protein antigens' ('antigens'). Antigen features contain information extracted from the descriptions in the SwissProt/TrEMBL (SPTR)^[18] or GenBank®^[3] entries. For each entry, the most relevant sequence is provided along with links to related entries from SPTR. The related antigen sequences have been compiled by searching SPTR with the source peptide sequence allowing no mismatch. Each related antigen contains one or more MHC-associated peptides that have FIMM entries. HLA (MHC) refers to HLA entries with official names. Related antigen infor-

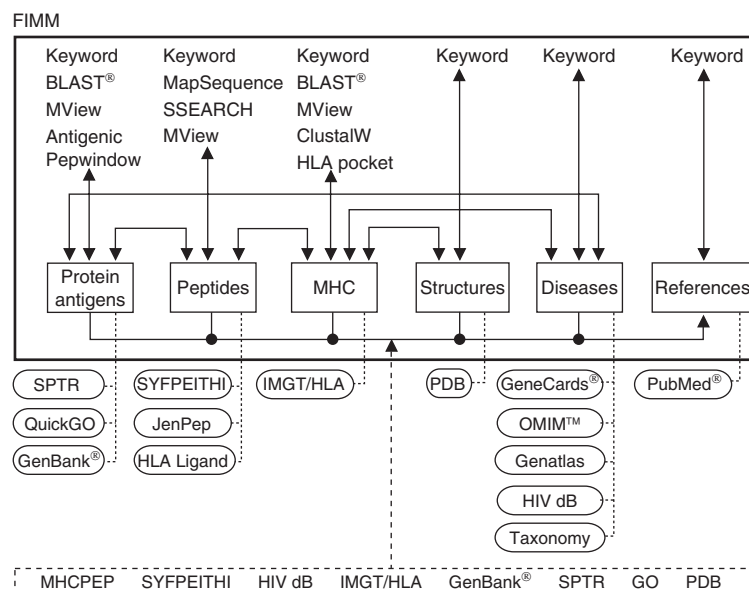


Fig. 1. Functional molecular IMMunology (FIMM) database data model. Solid lines with arrows indicate internal data connections. Dotted lines indicate hyperlinks to external data. Dashed lines indicate external data sources. **GO** = Gene Ontology™ database; **HLA** = human leukocyte antigens; **MHC** = major histocompatibility complex; **MHCPEP** = database of MHC binding peptides; **OMIM™** = Online Mendelian Inheritance in Man; **PDB** = Protein Data Bank; **SPTR** = SwissProt-TrEMBL database.

Table II. Functional molecular IMMunology (FIMM) database dimensions, data and growth. The number of peptide entries is higher than the number of sequences because peptides with identical sequences can bind to different human leukocyte antigens (HLA) molecules and were reported in different publications. Nonredundant peptides differ in their sequences, but may bind to the same HLA molecule. Peptides fall into the categories 'naturally processed' and/or a 'T-cell epitope' and/or a 'B-cell epitope'. Therefore, the sum of entries or sequence in these categories is not equal to the number of entries or sequences in 'Peptides'. The number of HLA structure entries is higher than the number of sequences because one HLA molecule can be complexed with different peptides

Dimensions	FIMM release (year)							
	1.3 (2003)		1.2 (2001)		1.1 (2000)		1.0 (1999)	
	sequences (n)	entries (n)	sequences (n)	entries (n)	sequences (n)	entries (n)	sequences (n)	entries (n)
Disease	NA	178	NA	117	NA	71	NA	71
disease in context of peptide	NA	126	NA	65	NA	52	NA	52
HLA-disease associations	NA	52	NA	52	NA	19	NA	19
Peptides	1 756	1 970	1 402	1 591	1 278	1 452	1 089	1 233
naturally processed	1 012	1 113	905	1 000	882	977	802	890
T-cell epitopes	584	674	577	668	457	534	328	380
T-cell epitopes naturally processed	83	86	80	83	60	63	40	42
B-cell epitopes	256	266	0	0	0	0	0	0
Protein antigens	692	692	571	571	524	524	463	463
related antigens	72 424	72 424	38 235	38 235	27 182	27 182	0	0
MHC (HLA)	1 589	1 589	1 136	1 136	875	875	875	875
Structures	577	627	501	541	0	0	0	0
HLA 3-D models without peptide	557	557	481	481	0	0	0	0
HLA 3-D from PDB	20	70	20	60	0	0	0	0
References	NA	3 766	NA	2 815	NA	2 478	NA	2 303

3-D = 3-dimensional; **HLA** = human leukocyte antigen; **MHC** = major histocompatibility complex; **NA** = not applicable; **PDB** = Protein Data Bank.

mation allows the user to evaluate antigens for potential epitope variation, cross-species conservation or mimicry.

This update of FIMM also contains 37 nonredundant protein allergens that are reported with the official World Health Organization/International Union of Immunological Societies (WHO/IUIS) nomenclature,^[19] if available. Gene Ontology™ terms^[20] describe the biological role of the antigen in terms of molecular function, cellular process or cellular component. The Gene Ontology™ terms were deduced from Swiss-Prot keywords using the spkw2go translation table (E. Camon, Gene Ontology™ Consortium, <http://www.geneontology.org/external2go/spkw2go>). The terms are linked to the QuickGO Gene Ontology™ browser (<http://www.ebi.ac.uk/ego>).

Major Histocompatibility Complex (MHC) View

Factual entries of MHC contain the common fields, the sequence and the publication sources. In addition, the 'MHC' dimension contains information on MHC-binding pocket composition,^[21] co-receptor (CD8) binding sites,^[22] TCR binding sites,^[23] and 3-D structures or homology models if available. Binding

pocket analysis allows users to select and inspect binding pockets for a subset of MHC alleles or for a specific phenotype. The major source of MHC sequence information for FIMM is the IMGT/HLA database. Since the MHC nomenclature for HLA alleles has changed from 4- to 6-digit allele names, we updated all records according to the new nomenclature rules. To our knowledge, only the FIMM database provides an up-to-date nomenclature for peptide-HLA associations.

Human Leukocyte Antigen Structures

FIMM contains 3-D structures of HLA molecules. The 3-D structures include entries extracted from the Protein Data Bank (PDB).^[24] FIMM also contains homology models of the majority of HLA-A, -B and -C structures. These predicted structure entries were generated based on a representative set of published class I MHC crystal structures available in the PDB. Protein models were built using the program Modeller^[25] and briefly energy minimised in a solvent bath using the molecular mechanics program AMBER.^[26] The 'HLA structure' view allows users to display and download available 3-D structures using the MDL® Chime viewer (<http://www.mdlchime.com/chime>). Users can display molecular

Table III. Distribution of antigens, peptides and epitopes by disease aetiology

Dimensions	Disease aetiology						
	neoplastic/ mutation	auto-immune	self-peptides/ -antigens	allergy	virus	microbe	parasite
Antigens	121	14	426	37	114	25	8
Related antigens	18 273	271	9475	3447	57 772	174	167
Peptides	289	74	1044	271	271	38	20
T-cell epitopes	225	12	246	12	252	34	19
B-cell epitopes	2	0	1	256	0	1	0

models in various representations, such as line model, ball-and-stick model or ribbon model. FIMM version 1.3 contains 70 PDB-derived entries of 20 HLA molecules complexed with peptides, β_2 microglobulin, TCR, CD8 or killer cell immunoglobulin-like receptors. Three-dimensional structures of 479 empty (without peptide) HLA-A, -B and -C molecules and 78 class II molecules were derived by homology modelling. Of 78 class II structures, 38 were modelled from DRB, 10 from DQB, 12 from DPB complete β chains and 12 from DRA, DQA, DPA complete α chains. Six structures are models of DQA/DQB and DPA/DPB heterodimers. Both PDB structures and models can be downloaded and imported into specialist programs for modelling peptide-MHC interactions.

Diseases

A factual entry from the dimension 'diseases' contains the following information: common fields, aetiology description (e.g. food allergen, virus, cancer) and MHC disease association (if applicable). Disease associations have been compiled from selected publications, such as HLA workshop proceedings or journal articles. Curators summarised the disease characteristics in context of the antigen or epitopes in the field 'description' and classified the entries by aetiology (table III). The majority of entries relate to cancers and viral infections, reflecting the importance of these areas in applied immunology. Self-peptides *per se* are not associated with a disease, but may contain epitopes of tumour- or autoantigen candidates. External links to the National Center for Biotechnology Information (NCBI) Taxonomy Browser,^[27] OMIMTM (Online Mendelian Inheritance in Man),^[28] GeneCards[®]^[29] and Genatlas^[30] databases, or selected World Health Organization (http://www.who.int/health_topics/en/) and Centers for Disease Control and Prevention (<http://www.cdc.gov/ncidod/>) websites, provide users with detailed disease information.

References

The 'reference' dimension contains publication references, namely the author names, title, journal and the links to PubMed[®] records. Users can search all references by keywords, author names or title words, or restrict the search target to references of

peptides, antigens or diseases. The retrieved references are hyperlinked to relevant data views of the FIMM database.

FIMM Query and Analysis Tools

FIMM contains a set of generic sequence similarity search tools and keyword search tools to perform specific queries and retrieve data from different views. Keyword searches with free-text input are available in each dimension except 'structure'. All keyword searches (both exact and partial matching) can be restricted in the advanced mode by user-defined categories (e.g. list HLA alleles by loci; diseases by aetiology; peptides by MHC molecules, T-cell or B-cell epitope). Extracted information is presented in the form of reports or lists containing hyperlinks to other views and external sources.

In this FIMM update, we improved the specificity of retrieval and the information value of the summary reports for the dimensions 'peptide', 'antigen' and 'disease'. Search results are reported in a table format according to search criteria, indicating whether the peptide sequence is a T-cell epitope, naturally processed HLA binder or a B-cell epitope. The keyword search function for 'antigen' includes Gene OntologyTM terms, permitting the retrieval of antigen and associated peptide and disease information by broader functional concepts.

FIMM integrates tools for sequence similarity search, regular expression search, multiple sequence alignment and protein antigen property analysis. Users can compare their query sequence with the FIMM entries using the BLAST2 program.^[31] A query sequence can be compared with MHC sequences, protein antigen or related antigen sequences from the SPTR database. The HLA alignment tool uses ClustalW^[32] for sequence alignment and MView^[33] for alignment colouring. The peptide search tool uses a local string-matching search program or the Smith-Waterman algorithm of SSEARCH^[34] customised for finding short sequences. These algorithms enable users to search the FIMM database for antigens that contain a subsequence identical (FIMM internal program MapSequence) or similar (SSEARCH) to a query peptide. The binding pocket tool displays the amino acids that form binding pockets in the groove of known HLA molecules. It also permits the alignment of a query MHC sequence to a known

HLA sequence and identification of pocket-forming positions in the query sequence of an MHC allele without experimental peptide binding data.

The addition of B-cell epitopes to FIMM necessitated the integration of two basic EMBOSS (European Molecular Biology Open Software Suite)^[35] programs for identifying epitopes within a protein sequence that may react or cross-react with an antibody recognising the native protein. Pepwindow displays a Kyte-Doolittle hydropathy plot^[36] of the antigen. The program Antigenic predicts potentially antigenic regions of a protein sequence, using the semi-empirical method of Kolaskar and Tongaonkar.^[37]

Conclusion

This update of FIMM focused on allergies, their associated IgE-binding B-cell epitopes, cancers and HLA class II structure models. Cross-checking of data for both accuracy and redundancy was performed and query functions were improved to serve the needs of the FIMM user community. However, there is a growing need for a comprehensive epitope-centred database that has broader information value for various subdisciplines of immunology. Therefore, we are planning a framework for integrating features of FIMM and related databases, for example SYFPEITHI or AntiJen, into a unified, nonredundant epitope resource without discontinuing the existing service of potential collaborating database providers.

Acknowledgements

We thank Dr Limsoon Wong (Institute for Infocomm Research) for support and discussion.

The authors declare that they have no competing financial interest.

References

- Marsh SG, Albert ED, Bodmer WF, et al. Nomenclature for factors of the HLA system, 2002. *Tissue Antigens* 2002; 60: 407-64
- Robinson J, Waller MJ, Parham P, et al. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res* 2003; 31: 311-4
- Benson DA, Karsch-Mizrachi I, Lipman DJ, et al. GenBank update. *Nucleic Acids Res* 2004; 32: D23-6
- Blythe MJ, Doytchinova IA, Flower DR. JenPep: a database of quantitative functional peptide data for immunology. *Bioinformatics* 2002; 18: 434-9
- Giudicelli V, Lefranc M-P. IMGT/GENE-DB. In: Galperin MY, editor. The Molecular Biology Database Collection: 2004 update [entry no. 503]. *Nucleic Acids Res* 2004; 32 (database issue). Available from URL: <http://www3.oup.co.uk/nar/database/summary/503> [Accessed 2005 May 9]
- Lefranc MP. IMGT, the international ImmunoGeneTics database. *Nucleic Acids Res* 2003; 31: 307-10
- Lefranc MP, Giudicelli V, Ginestoux C, et al. 2004. IMGT-ONTOLOGY for immunogenetics and immunoinformatics. *In Silico Biol* 2004; 4: 17-29
- Brusic V, Rudy G, Harrison LC. MHCPEP, a database of MHC-binding peptides: update 1997. *Nucleic Acids Res* 1998; 26: 368-71
- Rammensee H, Bachmann J, Emmerich NP, et al. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 1999; 50: 213-9
- Thorsby E. Invited anniversary review: HLA associated diseases. *Hum Immunol* 1997; 53: 1-11
- Brusic V, Millot M, Petrovsky N, et al. Allergen databases. *Allergy* 2003; 58: 1093-100
- Gendel SM. Sequence analysis for assessing potential allergenicity. *Ann N Y Acad Sci* 2002; 64: 87-98
- Charron D, editor. Proceedings of the Twelfth International Histocompatibility Workshop and Conference. Paris: Éditions Médicales et Scientifiques (EDK), 1997
- Ballard C, Herremans D, Schau D, et al. Data modeling techniques for data warehousing. San Jose: IBM International Support Organization, 1998
- Schönbach C, Koh JL, Flower DR, et al. FIMM, a database of functional molecular immunology: update 2002. *Nucleic Acids Res* 2002; 30: 226-9
- Korber BTM, Koup R, Walker BD, et al., editors. HIV molecular immunology 2002 [online]. Available from URL: <http://www.hiv.lanl.gov/content/immunology/pdf/2002/immuno2002.pdf> [Accessed 2005 May 9]
- Sathiamurthy M, Hickman HD, Cavett JW, et al. Population of the HLA ligand database. *Tissue Antigens* 2003; 61: 12-9
- Boeckmann B, Bairoch A, Apweiler R, et al. The SWISS-PROT protein knowledge base and its supplement TrEMBL in 2003. *Nucleic Acids Res* 2003; 31: 365-70
- King TP, Hoffman D, Lowenstein H, et al. Allergen nomenclature. *Allergy* 1995; 50: 765-74
- Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; 25: 25-9
- Chelvanayagam G. A roadmap for HLA-DR peptide binding specificities. *Hum Immunol* 1997; 58: 61-9
- Sullivan JA, Oettinger HF, Sachs DH, et al. Analysis of polymorphism in porcine MHC class I genes: alterations in signals recognized by human cytotoxic lymphocytes. *J Immunol* 1997; 159: 2318-26
- Hashimoto K, Okamura K, Yamaguchi H, et al. Conservation and diversification of MHC class I and its related molecules in vertebrates. *Immunol Rev* 1999; 167: 81-100
- Bourne PE, Address KJ, Bluhm WF, et al. The distribution and query systems of the RCSB Protein Data Bank. *Nucleic Acids Res* 2004; 32: D223-5
- Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993; 234: 779-815
- Cornell WD, Cieplak P, Bayly CI. A second generation force field for the simulation of proteins and nucleic acids. *J Am Chem Soc* 1995; 117: 5179-97
- Wheeler DL, Church DM, Edgar R, et al. Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Res* 2004; 32: D35-40
- Hamosh A, Scott AF, Amberger JS, et al. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 2002; 30: 52-5
- Rebhan M, Chalifa-Caspi V, Prilusky J, et al. GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics* 1998; 14: 656-64
- Frezal J. GenAtlas database, genes and development defects. *C R Acad Sci III* 1998; 321: 805-17
- Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; 25: 3389-402
- Chenna R, Sugawara H, Koike T, et al. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 2003; 31: 3497-500
- Brown NP, Leroy C, Sander C. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics* 1998; 14: 380-1
- Pearson WR. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. *Genomics* 1991; 11: 635-50
- Olson SA. EMBOSS opens up sequence analysis. *European Molecular Biology Open Software Suite. Brief Bioinform* 2002; 3: 87-91
- Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 1982; 157: 105-32
- Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinant on protein antigens. *FEBS Lett* 1990; 276: 172-4

Correspondence and offprints: Dr Christian Schönbach, Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613, Singapore. E-mail: fimm@i2r.a-star.edu.sg