

Special Feature

Computational immunology: The coming of age

NIKOLAI PETROVSKY¹ and VLADIMIR BRUSIC²

¹*National Bioinformatics Centre, University of Canberra and National Health Sciences Centre, Canberra Clinical School, Woden, Australian Capital Territory, Australia and* ²*Kent Ridge Digital Laboratories, Singapore*

Summary The explosive growth in biotechnology combined with major advances in information technology has the potential to radically transform immunology in the postgenomics era. Not only do we now have ready access to vast quantities of existing data, but new data with relevance to immunology are being accumulated at an exponential rate. Resources for computational immunology include biological databases and methods for data extraction, comparison, analysis and interpretation. Publicly accessible biological databases of relevance to immunologists number in the hundreds and are growing daily. The ability to efficiently extract and analyse information from these databases is vital for efficient immunology research. Most importantly, a new generation of computational immunology tools enables modelling of peptide transport by the transporter associated with antigen processing (TAP), modelling of antibody binding sites, identification of allergenic motifs and modelling of T-cell receptor serial triggering.

Key words: allergy, autoimmunity, bioinformatics, cancer, computational immunology, databases, genomics.

Introduction

The difficulties in the development and application of computational immunology tools include the high complexity of immunological processes, the intrinsic imprecision of biological measurements, and the biases and misconceptions inherent to the scientific endeavour. Care needs to be taken that computational tools are adequate and that they properly model the underlying immunological processes. Biases and misconceptions can easily be encoded in computational tools, resulting in the misinterpretation of insights provided by computational approaches. One role of computational immunology is to convert immunological into computational problems, solve computational problems using mathematical and computational approaches and convert these results into biologically meaningful interpretations.

Bioinformatic tools available for database searching and biological sequence analysis have become increasingly sophisticated. These tools enable the quick identification of sequences of interest and provide some bibliographic, taxonomic and feature information. Tools for sequence comparison, motif searching or sequence profiling assist researchers to identify biologically relevant sequence similarities and features. A newer generation of computational tools enables the modelling of biological interactions and simulation of laboratory experiments. These tools help researchers identify and focus upon relevant experiments, thus speeding up the discovery process. Therefore, biological databases, sequence comparison tools, and more complex modelling and simulation tools, are core resources for computer-assisted discovery and data analysis.

Correspondence: Associate Prof. N Petrovsky, National Health Sciences Centre, Autoimmunity Research Unit, Canberra Hospital, PO Box 11, Woden, ACT 2606, Australia.
Email: nikolai.petrovsky@anu.edu.au

Received 25 February 2002; accepted 25 February 2002.

Databases

There has been an explosion in general-purpose biological databases that contain annotated biological sequence entries.¹ Most commonly, these databases contain the actual sequence, a short description, the name and the taxonomy of the source organism, and in some cases tables listing features and bibliographic data. Amongst the best-known generic databases are the nucleic acid databases, GenBank² and European Molecular Biology Laboratory (EMBL),³ and the protein databases, SWISS-PROT⁴ and Protein Information Resource (PIR).⁵ Table 1 lists the URL addresses for a number of these generic databases. Information on protein families and domains, including information on biologically significant sites, patterns and profiles, can be accessed through the PROSITE database.⁶ This contains more than 10 000 3-D macromolecular structures, including several hundred structures of immune-system products.

Although the generic databases listed in Table 1 contain large numbers of sequences of immunological relevance, the annotation of the sequences in these databases is often limited to basic facts. Fortunately, more specific databases also exist that contain immunologically relevant sequences. A comprehensive list of databases is available through the Molecular Biology Database Collection,⁷ or online through the Public Catalogue of Databases (DBCAT) at <http://www.infobiogen.fr/services/dbcat/file/dbcat.html>. This site currently (January 2002) lists 87 DNA, 29 RNA, 94 protein, 58 genomic, 29 mapping, 18 protein structure, 43 literature and 153 miscellaneous databases.

Searching generic sequence databases is an essential tool in many immunology projects, providing useful hits or insights into a sequence of interest. When more sophisticated analyses are required (such as elucidation of related biological pathways or molecular function amongst others), more specific and detailed databases may be required. In many cases,

Table 1 A short list of the best-known, general-purpose sequence databases

Database	URL
DBCAT Catalogue of Databases	http://www.infobiogen.fr/services/dbcat
GenBank	http://www.ncbi.nlm.nih.gov
EMBL	http://www.ebi.ac.uk/embl
Protein Information Resource (PIR)	http://www-nbrf.georgetown.edu/pir
SWISS-PROT	http://www.expasy.ch/sprot
PROSITE	http://www.expasy.ch/prosite
PDB	http://www.rcsb.org/pdb

EMBL, European Molecular Biology Laboratory; PDB, Protein data bank.

these specialty databases provide better overall quality of information, at least partly because they are annotated by experts from that particular field. This raises the commonly neglected issue of database quality that has major relevance for casual users of both generic and specialty databases alike and will be dealt with later in this review.

Specialist immunology databases

An example of a specialist immunology database is the HIV Molecular Immunology Database (<http://hiv-web.lanl.gov/immunology/>). This contains T-cell epitope maps on HIV proteins, alignments and annotations, and links to the HIV Sequence Database (<http://hiv-web.lanl.gov/>) that contains HIV genetic sequences, immunological epitopes, drug-resistance-associated mutations and vaccine trials. Annotation includes information such as how specific epitopes were experimentally defined, HLA specificities, isotypes of mAb, the initial immunogen, and brief notes describing the context in which a given epitope was studied. Associated database tools include Epilign (aligns amino acid epitopes against alignments), PeptGen (maps overlapping peptides, with options to adjust length, overlap, and to exclude user-selected amino acids from C- or N-terminal positions), Motifscan (scans sequences for possible epitopes based on HLA-binding motifs), Primalign (aligns nucleotide sequences against alignments), the HXB2 numbering engine, SeqPublish and BLAST.

The International ImMunoGeneTics database (IMGT; <http://imgt.cines.fr:8104/>)⁸ is a high-quality integrated database specializing in immunoglobulins, T-cell receptors, and MHC of all vertebrate species. At present, IMGT includes two databases: (i) LIGM-DB, a comprehensive database of immunoglobulins and T-cell receptors from human and other vertebrates, with translation for fully annotated sequences; and (ii) HLA-DB, a database of HLA.

KABAT (<http://immuno.bme.nwu.edu/>)⁹ the oldest specialist immunology database, includes nucleotide sequences, sequences of TCR, MHC class I and II molecules, and other proteins of immunological interest. It includes tools such as SeqhuntII, Variability, Align-A-Sequence, Subgrouping and Find Your Family.

The Swine Major Histocompatibility Complex (SLA) database (SLAD; <http://sdmc.krdl.org.sg/bic/projects/p19.html>) provides the basis for establishing SLA nomenclature through phylogenetic analyses.

MHCPEP (<http://wehih.wehi.edu.au/mhcpep/>)¹⁰ is a database comprising over 13 000 peptide sequences known to bind MHC molecules. Entries were compiled from published reports as well as from direct submissions of experimental data. Each entry contains the peptide sequence, its MHC specificity and, when available, experimental method, observed activity, binding affinity, source protein, anchor positions and publication references.

FIMM is an annotated database of functional immunology.¹¹ FIMM includes data on MHC, protein antigens, antigenic peptides and relevant disease information. Tools include searches by key word, sequence pattern and basic local alignment search tool (BLAST), multiple sequence alignment and binding pocket/contact sites analysis.

The SYFPEITHI database (<http://syfpeithi.bmi-heidelberg.com/scripts/MHCServer.dll/home.htm>) contains data on peptide sequences, anchor positions, MHC specificity and tools for epitope prediction.¹²

FARRP (<http://www.allergenonline.com/default.asp>) contains a list of publicly known allergens. The National Center for Food Safety and Technology has created three databases of known allergen sequences in order to facilitate the use of sequence comparisons in assessing the potential allergenicity of food proteins (<http://www.iit.edu/~sgendel/fa.htm>). These databases contain accession numbers for food allergens, non-food allergens and wheat gluten proteins.

The Protall Database (<http://www.ifrn.bbsrc.ac.uk/protall/database.html>) lists biochemical and clinical information about plant food allergens involved in classical IgE-induced hypersensitivity reactions.

Prediction Algorithm for Proteasomal Cleavages (PAPROC; <http://www.paproc.de>), whilst not strictly a database, is a potentially useful immunological tool. PAPROC predicts cleavage by human and yeast proteasomes, based on experimental cleavage data and is thus useful for immunologists working on antigen processing and the prediction of MHC class I molecule ligands and CTL epitopes.¹³

Bioinformatics & Molecular Analysis Section (BIMAS; <http://bimas.dcrn.nih.gov/>) is a site designed to provide guidance, support and resources to scientists throughout the National Institutes of Health in bioinformatics, genomics and genetic analysis. Examples of BIMAS tools accessible over the web include an HLA-binding prediction program, promoter scan (the PROSCAN program), a signal scan (SIGSCAN), sequence format conversion program (READSEQ) and identification of tissue-enriched (human or mouse) unigene clusters.

The specialist immunology databases are becoming increasingly numerous and sophisticated. The functionality of databases will increase as more high-level analysis tools are integrated into databases. However, this additional complexity will also require increasingly sophisticated and knowledgeable users if traps and erroneous results are to be avoided. Some of the potential traps arising from database design and usage are described below.

Database issues

Data standardization

Standardization and nomenclature are important so as to allow a consistent view of the data available in a particular

field. Databases play an important role because of the need to unify existing scientific classification and nomenclature with data structures suitable for analysis by computers. An important objective of the IMGT database is to contribute to the standardization of data in immunology. Amongst other uses, standardized nomenclature enables cross-species comparisons, thereby providing avenues to gain new insights into molecular structure/function relationships.

Data interpretation

Database developers are well aware of the difficulties in building reliable and effective databases. However, if users are to efficiently use these resources then they also need a good understanding of relevant issues. The interpretation of data extracted from databases is highly dependent on the skills and knowledge of the user. Lack of standards, ad hoc nomenclature, variable source data quality, incomplete information and biases are all major obstacles to data interpretation. Furthermore, most database search tools require careful selection of parameters for search optimization, as default settings are often not optimal for a specific query. The selection of search parameters requires a good understanding of the specific issues associated with a particular query. The development of specialist databases enhances the quality and availability of data, but human experts are still necessary for annotation and interpretation.

Data quality

Database quality is a descriptive measure of the extent to which a database is capable of achieving its intended purpose. As researchers and end-users, we would like to think that all the data held within a database is of uniformly high quality. Unfortunately, this is rarely the case. In order to properly assess database quality we need to address its individual components. These include data accuracy and accessibility, ease of use, breadth and depth of coverage, ease of integration with data from other sources, and performance of data analysis tools attached to the database. Errors in public databases can result from biological variation, experimental artefacts, imperfect sequence determination and poor data handling, amongst others. For example, the GenBank database included 0.36% vector-contaminated sequences in release 1995–1996 and had a 10–20% rate of erroneous annotation of entries as 'genes'.¹⁴ Although current standards allow for genomic sequences to have an error rate of less than one error per 10000 bases,¹⁵ the sequence error rate is much higher than this for older data. This has been compounded by the use of automatic annotation of database entries, which, although faster, tends to expand existing errors and introduce new errors. Overall error rates in functional assignment may be as high as 2.5–5.0%. What this demonstrates is the need for careful annotation, including human expert assessment of each database entry. Examples of expert-annotated immunological databases include the KABAT,⁹ IMGT⁸ and FIMM¹¹ databases.

Database tools

Databases require tools to enable efficient extraction of data and generation of knowledge. Therefore, to increase their

research efficiency, immunologists require both efficient database facilities to manage and store complex information plus relevant data, and sequence analysis tools to perform complex analyses on extracted data. Hundreds of programs for biological data analysis have been developed in recent years. A listing of more than 130 of these bioinformatics tools is accessible via eBioinformatics BioNavigator at <http://www.bionavigator.com>. The FIMM database,¹¹ for example, integrates BLAST and pattern searches, multiple sequence analysis and binding site analysis in the database interface. Databases that integrate tools enabling complex data analysis will be increasingly indispensable to biomedical research. Although individual databases are often readily accessible, it is often difficult to extract information from heterogeneous databases because of a lack of standard nomenclature and variable formats. This problem is slowly being overcome and systems such as Kleisli¹⁶ or CORBA¹⁷ provide means for efficient extraction of data from heterogeneous sources. These systems facilitate the automation of large-scale data access, extraction and integration.

Specific applications of computational immunology

Allergies

The number of known protein allergens is increasing rapidly, with some 360 protein allergens having been classified as of September 2001. The assessment of protein allergenic potential focuses on three main aspects: (i) immunogenicity; (ii) cross-reactivity; and (iii) clinical symptoms.¹⁸ Immunogenicity reflects the likelihood of an IgE antibody and/or T-cell response to a particular allergen. Therefore, immunogenicity studies focus on identifying B-cell^{19,20} and T-cell^{21,22} recognition sites on allergens. The 3-D structural properties of allergens control their allergenicity, and four families of allergens have been defined recently based on 3-D structure.²³ Allergen-related data are currently spread across a number of publicly accessible molecular databases, including the major protein databases (PIR, SwissProt and TrEMBL). As these databases do not contain detailed structural and functional allergen information, various specialized allergy-related databases and data repositories have been developed. Bioinformatic prediction tools enabling the identification of proteins with allergenic potential are available for prediction of allergenicity, allergen cross-reactivity,²⁴ and T-cell epitopes.²² Structural bioinformatic tools help identify structural properties of proteins, such as secondary structure,²⁵ or tertiary structure,^{26,27} which affect allergenicity through affecting IgE binding sites. Through the collation and analysis of the protein sequences of known allergens and their epitopes, it is possible to identify related groups that correlate with observed clinical cross-reactivities. Comparative molecular modelling was successful in identifying a common putative IgE epitope on cysteine protease allergens of diverse sources.²³

The use of bioinformatic tools to predict protein allergenicity will become increasingly important in the screening of novel foods before their wide-scale release for human use.¹⁸ Thus, there is a major need to develop reliable broad-based allergy databases and combine these with well-validated bioinformatic prediction tools in order to enable the identification of potential allergens in genetically modified drugs and foods.

Infectious diseases and host responses

Mathematical and computer modelling can be of great help in understanding the behaviour and spread of infectious disease. Examples include mathematical modelling of the within-host dynamics of *Plasmodium falciparum*,¹⁹ determination of host factors influencing viral persistence,²⁰ development of a stochastic model of schistosomiasis immuno-epidemiology,²¹ and of nematode infection in ruminants,²² modelling the dynamics of lymphocytic choriomeningitis virus (LCMV) infection in mice,²⁸ and modelling the likelihood of rubella epidemics in Europe.²⁹ Computer simulations have been performed on the transmission dynamics and the effects of duration of immunity and survival of persistently infected animals in the spread of bovine viral diarrhoea virus in dairy cattle.³⁰ *Trypanosoma congolense* parasitaemia patterns have been modelled during the chronic phase of infection in N'Dama cattle.³¹ In addition, there has been development of a theoretical framework for the immunoepidemiology of *Plasmodium falciparum* malaria,³² and of the development resistance of *P. falciparum* to antimalarial drugs.³³ More recently, extensive work has been conducted on modelling HIV vaccination,³⁴ and HIV-infection therapy.³⁵

A modelling approach has also been used to better understand control of infections by immunization. This has been studied in respect of rabies control in fox populations,³⁶ the modelling of antibody response to measles vaccine and subsequent waning of immunity in a low exposure population,³⁷ the impact of immunization on the epidemiology of Varicella zoster virus,³⁸ and the impact of mass vaccination against hepatitis B.³⁹ Mathematical models have been used to assess the long-term persistence of antibodies after vaccination with two inactivated hepatitis A vaccines,^{40,41} and to identify subcritical endemic steady states for animal infections with incomplete immunity.²⁴

These examples illustrate the power of a computational approach to complex problems involving potentially vast datasets of sometimes imprecise clinical data. The development of computer models of infectious diseases and their response to treatment or vaccination allows the initial testing of hypotheses, which can then be validated in the field or the laboratory. This offers the ability to speed up the discovery process dramatically and to potentially shorten the time for vaccine development for major infections such as HIV.

Immune system function

Modelling has also been extremely helpful in the understanding of immune population dynamics. It has been used to model T-cell-mediated suppression,²⁵ peripheral lymphocyte migration,²⁶ T-cell memory,^{27,42} tolerance,^{43,44} thymic function,^{45,46} and antibody networks.⁴⁷ Models predict the dynamics of antiviral cytotoxic T-cell memory in response to different stimuli.⁴⁸ Shape-space modelling has been successfully used to understand the nature of specificity in an immune network.⁴⁹ Modelling has been particularly successful in predicting TCR serial triggering and downregulation,⁵⁰ HLA binding,^{51,52} T-cell epitopes,⁵³ and immunogenicity.^{54,55}

A demonstration of the power of a computational modelling approach to the understanding of immune system function is shown by a recent study where we sought to examine the function relationship between TAP peptide transport and

HLA class I antigen presentation.⁵⁶ TAP is a transmembrane protein responsible for the transport of antigenic peptides into the endoplasmic reticulum, where they can be bound by MHC class I molecules and presented to T cells. As TAP does not bind all peptides equally, TAP-binding affinity could influence the ability of a particular peptide to gain access to the MHC class I pathway. However, analysis of the relationship between TAP- and MHC class I-binding has been hampered by the extensive polymorphism of MHC molecules (>200 HLA class I allelic variants exist). To overcome these problems, we used an artificial neural network (ANN) computer model to study peptide binding to human TAP and its relationship with MHC class I binding.⁵⁶ After development and validation, the ANN model of TAP-peptide binding was used to mine a database of HLA-binding peptides to elucidate patterns of TAP binding. The affinity of HLA-binding peptides for TAP was found to differ according to the HLA supertype concerned: HLA-B27-, -A3- or -A24-binding peptides had high affinities for TAP, whereas HLA-A2-, -B7- or -B8-binding peptides had low affinities for TAP. The affinities of HLA-A2- and HLA-B7-binding peptides for TAP were shown to have similar distributions to that of randomly generated peptides. This work showed for the first time that HLA molecules constitute two separate classes: (i) those that are TAP-dependent for peptide loading (HLA-B27, -A3 and -A24); and (ii) those that are relatively TAP-independent (HLA-A2, -B7 and -B8). The strong similarity between the sets of peptides bound by TAP and HLA-B27, -A3 or -A24 suggest the possibility of their functional coevolution. This could have important implications for the design of peptide-based immuno-therapeutic drugs and vaccines, and shows the power of the modelling approach in helping to better understand complex immune interactions.

Antibody-antigen-MHC interactions

Various models have been developed to understand antibody-antigen interactions. Modelling of ferritin antibody-antigen interactions based on experimental data helped explain the inability of the antibody to recognize human L-ferritin and rat, chicken and mouse H-ferritin.⁵⁷ Models have been used to study IgE/receptor interactions,⁵⁸ and the MHC class II allele I-A(g7) of NOD mice.⁵⁹ A number of studies have shown the power of combining experimental results with a modelling approach to understand antibody-antigen interactions.^{57,60} Other examples of the use of immune interaction modelling include the prediction of epitopes and the production of mAb against gastric H,K-ATPase,⁶¹ modelling of the 3-D structure for EgDf1 from *Echinococcus granulosus*,⁶² modelling the kinetics of antigen-antibody reactions in particle-enhanced optical immunoassays,⁶³ the construction of an antibody against cystatin,⁶⁴ the use of cellular automaton to study local T-T-cell and T-B-cell interactions,⁶⁵ and to demonstrate the importance of efficacy and partial agonism in models of B-lymphocyte activation.⁶⁶

These examples demonstrate the power of modelling approaches in the understanding of the interaction between complex 3-D surfaces, such as the binding of antibody to antigen. In addition, computational approaches are also well suited to numerical problems, such as understanding the

multiplicity of antibody and T-cell idiotypes and how these all come together to constitute an immune response.

Autoimmunity

Modelling has mainly been used in the autoimmunity area for the prediction of HLA-binding peptides and T-cell epitopes.⁵³ Other uses have included modelling of MHC susceptibility to rheumatoid arthritis,⁶⁷ and identification of antibody epitopes within the CB-11 peptide of type II collagen.⁶⁸ Molecular modelling of HLA-DQ has also been used to suggest a mechanism of resistance in type 1 diabetes.⁶⁹ More general models of autoimmunity and its therapy,⁷⁰ and the effects of genetic and environmental factors,⁷¹ have been developed.

Given the basis of autoimmunity being one of the classic unsolved philosophical problems of modern immunology, the scarcity of computational models that have been developed to try and resolve this problem is surprising. Similarly, there is a scarcity of databases with a major focus on autoimmunity and its mechanisms. This is something that we hope to see addressed in the near future.

Tumour immunology

Computer models have been used in an attempt to understand the processes of adoptive cellular cancer immunotherapy.⁷² Macrophage invasion of tumours and their effects on growth and composition have also been modelled.⁷³ Predictions have been used successfully to identify tumour-derived peptides suitable for use in cancer vaccines.^{74–76}

This is a relatively new field; therefore, the bioinformatic resources for tumour immunology research remain limited. However, given the increasing success of tumour vaccination, these resources are likely to grow significantly in the near future.

Organ transplantation

Neural networks have been used to predict the outcome of liver,^{77,78} pancreas,³³ and bone marrow transplantation.⁷⁹ Other examples of the use of predictive models in transplantation include the prediction of tacrolimus blood levels in liver transplantation,³⁰ diagnosis of early acute renal allograft rejection and evaluation of complications of renal transplants,³¹ prediction of cytomegalovirus disease after renal transplantation,³² and prediction of delayed renal allograft function.³⁶

As an example of how a computational prediction system can assist in improving transplantation outcomes, we recently showed that a computational prediction system can accurately predict graft rejection for particular donor/recipient pairs.⁸⁰ The prediction system used an ANN trained on renal transplant data from the Australian and New Zealand Dialysis and Transplant Registry (ANZDATA). After training with data from 1400 renal transplant recipients, the ANN correctly predicted 84.95% of successful and 71.7% of unsuccessful, transplants. We then tested whether an ANN could predict the type of graft rejection, namely hyper-acute, acute, subacute and chronic. The ANN correctly predicted rejection type in 59% of cases compared to a random probability of 25%. These results indicated that an ANN-based approach is useful

for predicting both the overall outcomes of renal transplants as well as the type of rejection. Conceivably, computational prediction systems have great potential as tools for cadaver kidney allocation, management of post-transplantation immunosuppression and the exploration of the role of HLA in transplant rejection.

Conclusions

Computational immunology is a rapidly expanding field that has moved far beyond its origins as a series of databases and search tools. More and more advanced modelling technologies are enabling laboratory scientists to better understand their data and to develop new hypotheses to explain immune system function. The power of this combined approach is most evident in the close correlation between wet lab experimental data and the results obtained using computational models.

References

- 1 Brusic V, Zeleznikow J, Petrovsky N. Molecular immunology databases and data repositories. *J. Immunol. Meth.* 2000; **238**: 17–28.
- 2 Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL. GenBank. *Nucleic Acids Res.* 2002; **30**: 17–20.
- 3 Stoesser G, Baker W, van den Broek A *et al.* The EMBL nucleotide sequence database. *Nucleic Acids Res.* 2002; **30**: 21–6.
- 4 Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 2000; **28**: 45–8.
- 5 Wu CH, Huang H, Arminski L *et al.* The Protein Information Resource: an integrated public resource of functional annotation of proteins. *Nucleic Acids Res.* 2002; **30**: 35–7.
- 6 Falquet L, Pagni M, Bucher P *et al.* The PROSITE database, its status in 2002. *Nucleic Acids Res.* 2002; **30**: 235–8.
- 7 Baxevanis AD. The Molecular Biology Database Collection: 2002 update. *Nucleic Acids Res.* 2002; **30**: 1–12.
- 8 Robinson J, Malik A, Parham P, Bodmer JG, Marsh SG. IMGT/HLA database – a sequence database for the human major histocompatibility complex. *Tissue Antigens* 2000; **55**: 280–7.
- 9 Johnson G, Wu TT. Kabat Database and its applications: future directions. *Nucleic Acids Res.* 2001; **29**: 205–6.
- 10 Brusic V, Rudy G, Harrison LC. MHCPEP, a database of MHC-binding peptides: update 1997. *Nucleic Acids Res.* 1998; **26**: 368–71.
- 11 Schonbach C, Koh JL, Flower DR, Wong L, Brusic V. FIMM, a database of functional molecular immunology: update 2002. *Nucleic Acids Res.* 2002; **30**: 226–9.
- 12 Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI database for MHC ligands and peptide motifs. *Immunogenetics* 1999; **50**: 213–19.
- 13 Nussbaum AK, Kuttler C, Haderl KP, Rammensee HG, Schild H. PAPROC: a prediction algorithm for proteasomal cleavages available on the WWW. *Immunogenetics*. 2001; **53**: 87–94.
- 14 Seluja GA, Farmer A, McLeod M, Harger C, Schad PA. Establishing a method of vector contamination identification in database sequences. *Bioinformatics* 1999; **15**: 106–10.
- 15 Macauley J, Wang H, Goodman N. A model system for studying the integration of molecular biology databases. *Bioinformatics* 1998; **14**: 575–82.
- 16 Chung SY, Wong L. Kleisli: a new tool for data integration in biology. *Trends Biotechnol.* 1999; **17**: 351–5.

- 17 Stevens R, Miller C. Wrapping and interoperating bioinformatics resources using CORBA. *Brief Bioinform.* 2000; **1**: 9–21.
- 18 Oehlschlager S, Reece P, Brown A *et al.* Food allergy – towards predictive testing for novel foods. *Food Addit. Contam.* 2001; **18**: 1099–107.
- 19 Hoshen MB, Heinrich R, Stein WD, Ginsburg H. Mathematical modelling of the within-host dynamics of *Plasmodium falciparum*. *Parasitology* 2000; **121**: 227–35.
- 20 Thomsen AR, Nansen A, Andreasen SO, Wodarz D, Christensen JP. Host factors influencing viral persistence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2000; **355**: 1031–41.
- 21 Chan MS, Isham VS. A stochastic model of schistosomiasis immuno-epidemiology. *Math. Biosci.* 1998; **151**: 179–98.
- 22 Marion G, Renshaw E, Gibson G. Stochastic effects in a model of nematode infection in ruminants. *IMA J. Math. Appl. Med. Biol.* 1998; **15**: 97–116.
- 23 Furmonaviciene R, Sewell HF, Shakib F. Comparative molecular modelling identifies a common putative IgE epitope on cysteine protease allergens of diverse sources. *Clin. Exp. Allergy* 2000; **30**: 1307–13.
- 24 Greenhalgh D, Diekmann O, de Jong MC. Subcritical endemic steady states in mathematical models for animal infections with incomplete immunity. *Math. Biosci.* 2000; **165**: 1–25.
- 25 Leon K, Perez R, Lage A, Carneiro J. Modelling T-cell-mediated suppression dependent on interactions in multicellular conjugates. *J. Theor. Biol.* 2000; **207**: 231–54.
- 26 Srikusalanukul W, De Bruyne F, McCullagh P. Modelling of peripheral lymphocyte migration: system identification approach. *Immunol. Cell Biol.* 2000; **78**: 288–93.
- 27 Jacob J, Baltimore D. Modelling T-cell memory by genetic marking of memory T cells *in vivo*. *Nature* 1999; **399**: 593–7.
- 28 Bocharov GA. Modelling the dynamics of LCMV infection in mice: conventional and exhaustive CTL responses. *J. Theor. Biol.* 1998; **192**: 283–308.
- 29 Edmunds WJ, van de Heijden OG, Eerola M, Gay NJ. Modelling rubella in Europe. *Epidemiol. Infect.* 2000; **125**: 617–34.
- 30 Innocent G, Morrison I, Brownlie J, Gettinby G. A computer simulation of the transmission dynamics and the effects of duration of immunity and survival of persistently infected animals on the spread of bovine viral diarrhoea virus in dairy cattle. *Epidemiol. Infect.* 1997; **119**: 91–100.
- 31 Agur Z, Mehr R. Modelling *Trypanosoma congolense* parasitaemia patterns during the chronic phase of infection in N'Dama cattle. *Parasite Immunol.* 1997; **19**: 171–82.
- 32 Gupta S, Day KP. A theoretical framework for the immuno-epidemiology of *Plasmodium falciparum* malaria. *Parasite Immunol.* 1994; **16**: 361–70.
- 33 Cross AP, Singer B. Modelling the development of resistance of *Plasmodium falciparum* to anti-malarial drugs. *Trans. R. Soc. Trop. Med. Hyg.* 1991; **85**: 349–55.
- 34 McLean AR, Blower SM. Modelling HIV vaccination. *Trends Microbiol.* 1995; **3**: 458–62.
- 35 Hrabá T, Dolezal J. Mathematical modelling of HIV infection therapy. *Int. J. Immunopharmacol.* 1995; **17**: 523–6.
- 36 Suppo C, Naulin JM, Langlais M, Artois M. A modelling approach to vaccination and contraception programmes for rabies control in fox populations. *Proc. R. Soc. Lond. B Biol. Sci.* 2000; **267**: 1575–82.
- 37 Mossong J, O'Callaghan CJ, Ratnam S. Modelling antibody response to measles vaccine and subsequent waning of immunity in a low exposure population. *Vaccine* 2000; **19**: 523–9.
- 38 Brisson M, Edmunds WJ, Gay NJ, Law B, De Serres G. Modelling the impact of immunization on the epidemiology of varicella zoster virus. *Epidemiol. Infect.* 2000; **125**: 651–69.
- 39 McLean AR, Blumberg BS. Modelling the impact of mass vaccination against hepatitis B. I. Model formulation and parameter estimation. *Proc. R. Soc. Lond. B Biol. Sci.* 1994; **256**: 7–15.
- 40 Van Herck K, Beutels P, Van Damme P *et al.* Mathematical models for assessment of long-term persistence of antibodies after vaccination with two inactivated hepatitis A vaccines. *J. Med. Virol.* 2000; **60**: 1–7.
- 41 Vellinga A, Van Damme P, Bruckers L, Weyler JJ, Molenberghs G, Meheus A. Modelling long-term persistence of hepatitis B antibodies after vaccination. *J. Med. Virol.* 1999; **57**: 100–3.
- 42 McLean AR. Modelling T cell memory. *J. Theor. Biol.* 1994; **170**: 63–74.
- 43 Dolezal J, Hrabá T. A contribution to mathematical modelling of immunological tolerance. *Arch. Immunol. Ther. Exp. (Warsz)* 1988; **36**: 23–30.
- 44 Borghans JA, de Boer RJ. Neonatal tolerance revisited by mathematical modelling. *Scand. J. Immunol.* 1998; **48**: 283–5.
- 45 Mehr R, Segel L, Sharp A, Globerson A. Colonization of the thymus by T cell progenitors: models for cell–cell interactions. *J. Theor. Biol.* 1994; **170**: 247–57.
- 46 Morpurgo D, Serenita R, Seiden PE, Celada F. Modelling thymic functions in a cellular automaton. *Int. Immunol.* 1995; **7**: 505–16.
- 47 Faro J, Carneiro J, Velasco S. Further studies on the problem of immune network modelling. *J. Theor. Biol.* 1997; **184**: 405–21.
- 48 Bocharov G, Klenerman P, Ehl S. Predicting the dynamics of antiviral cytotoxic T-cell memory in response to different stimuli: cell population structure and protective function. *Immunol. Cell Biol.* 2001; **79**: 74–86.
- 49 Harada K, Ikegami T. Evolution of specificity in an immune network. *J. Theor. Biol.* 2000; **203**: 439–49.
- 50 Sousa J, Carneiro J. A mathematical analysis of TCR serial triggering and down-regulation. *Eur. J. Immunol.* 2000; **30**: 3219–27.
- 51 Brusica V, Rudy G, Honeyman G, Hammer J, Harrison L. Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics* 1998; **14**: 121–30.
- 52 Lim JS, Kim S, Lee HG, Lee KY, Kwon TJ, Kim K. Selection of peptides that bind to the HLA-A2.1 molecule by molecular modelling. *Mol. Immunol.* 1996; **33**: 221–30.
- 53 Honeyman MC, Brusica V, Harrison LC. Strategies for identifying and predicting islet autoantigen T-cell epitopes in insulin-dependent diabetes mellitus. *Ann. Med.* 1997; **29**: 401–4.
- 54 Brusica V, Bucci K, Schonbach C, Petrovsky N, Zeleznikow J, Kazura JW. Efficient discovery of immune response targets by cyclical refinement of QSAR models of peptide binding. *J. Mol. Graph. Model.* 2001; **19**: 467.
- 55 Forster MJ, Rijpkema SG. Selection of putative immunogenic peptides by molecular modelling of the urease of *Helicobacter pylori*. *Dev. Biol. (Basel)* 2000; **103**: 75–9.
- 56 Brusica V, van Endert P, Zeleznikow J, Daniel S, Hammer J, Petrovsky N. A neural network model approach to the study of human TAP transporter. *In Silico Biol.* 1999; **1**: 109–21.
- 57 Helmer-Citterich M, Rovida E, Luzzago A, Tramontano A. Modelling antibody–antigen interactions: ferritin as a case study. *Mol. Immunol.* 1995; **32**: 1001–10.
- 58 Padlan EA, Helm BA. Modelling study of IgE/receptor interactions. *Biochem. Soc. Trans.* 1993; **21**: 963–7.
- 59 Moustakas AK, Routsias J, Papadopoulos GK. Modelling of the MHC II allele I-A (g7) of NOD mouse: pH-dependent changes in specificity at pockets 9 and 6 explain several of the unique properties of this molecule. *Diabetologia* 2000; **43**: 609–24.

- 60 Tenette-Souaille C, Smith JC. Structure of the Malpha2-3 toxin alpha antibody-antigen complex: combination of modelling with functional mapping experimental results. *Protein Eng.* 2000; **13**: 345-51.
- 61 Imaten M, Gallet X, Festy F *et al.* Prediction of epitopes and production of mAb against gastric H,K-ATPase. *Protein Eng.* 1998; **11**: 949-55.
- 62 Paulino M, Esteves A, Vega M, Tabares G, Ehrlich R, Tapia O. Modelling a 3D structure for EgDf1 from *Echinococcus granulosus*. Putative epitopes, phosphorylation motifs and ligand. *J. Comput. Aided Mol. Des.* 1998; **12**: 351-60.
- 63 Quesada M, Puig J, Delgado JM, Hidalgo-Alvarez R. Modelling the kinetics of antigen-antibody reactions at particle enhanced optical immunoassays. *J. Biomater. Sci. Polym. Ed.* 1998; **9**: 961-71.
- 64 Schiweck W, Skerra A. The rational construction of an antibody against cystatin: lessons from the crystal structure of an artificial Fab fragment. *J. Mol. Biol.* 1997; **268**: 934-51.
- 65 Hogeweg P. Local T-T cell and T-B cell interactions: a cellular automaton approach. *Immunol. Lett.* 1989; **22**: 113-22.
- 66 Hodgkin PD, Chin SH, Bartell G *et al.* The importance of efficacy and partial agonism in evaluating models of B lymphocyte activation. *Int. Rev. Immunol.* 1997; **15**: 101-27.
- 67 Genin E, Babron MC, McDermott MF *et al.* Modelling the major histocompatibility complex susceptibility to RA using the MASC method. *Genet. Epidemiol.* 1998; **15**: 419-30.
- 68 Brass A, Worthington J, Chen Y, Morgan K. Identification of antibody epitopes within the CB-11 peptide of type II collagen. II. Computer modelling studies of peptides and the interpretation of epitope scanning results. *Autoimmunity* 1991; **10**: 209-16.
- 69 Hoover ML, Marta RT. Molecular modelling of HLA-DQ suggests a mechanism of resistance in type 1 diabetes. *Scand. J. Immunol.* 1997; **45**: 193-202.
- 70 Waniewski J, Prikrylova D. Autoimmunity and its therapy: mathematical modelling. *Immunol. Lett.* 1988; **18**: 77-80.
- 71 Baxter AG. Modelling the effects of genetic and environmental factors on the risk of autoimmune disease. *J. Autoimmun.* 2001; **16**: 331-5.
- 72 Nani FK, Oguztoreli MN. Modelling and simulation of Rosenberg-type adoptive cellular immunotherapy. *IMA J. Math. Appl. Med. Biol.* 1994; **11**: 107-47.
- 73 Owen MR, Sherratt JA. Modelling the macrophage invasion of tumours: effects on growth and composition. *IMA J. Math. Appl. Med. Biol.* 1998; **15**: 165-85.
- 74 Zarour HM, Maillere B, Brusic V *et al.* NY-ESO-1 119-143 is a promiscuous major histocompatibility complex class II T-helper epitope recognized by Th1- and Th2-type tumor-reactive CD4+ T cells. *Cancer Res.* 2002; **62**: 213-18.
- 75 Zarour HM, Storkus WJ, Brusic V, Williams E, Kirkwood JM. NY-ESO-1 encodes DRB1*0401-restricted epitopes recognized by melanoma-reactive CD4+ T cells. *Cancer Res.* 2000; **60**: 4946-52.
- 76 Zarour HM, Kirkwood JM, Kierstead LS *et al.* Melan-A/MART-1 (51-73) represents an immunogenic HLA-DR4-restricted epitope recognized by melanoma-reactive CD4 (+) T cells. *Proc. Natl Acad. Sci. USA* 2000; **97**: 400-5.
- 77 Dvorchik I, Subotin M, Marsh W, McMichael J, Fung JJ. Performance of multi-layer feedforward neural networks to predict liver transplantation outcome. *Meth. Inf. Med.* 1996; **35**: 12-18.
- 78 Parmanto B, Doyle HR. Recurrent neural networks for predicting outcomes after liver transplantation: representing temporal sequence of clinical observations. *Meth. Inf. Med.* 2001; **40**: 386-91.
- 79 Keiding N, Klein JP, Horowitz MM. Multi-state models and outcome prediction in bone marrow transplantation. *Stat. Med.* 2001; **20**: 1871-85.
- 80 Petrovsky N, Tam S, Brusic V, Russ G, Socha L, Bajic V. Use of artificial neural networks in improving renal transplantation outcomes. *Graft* 2002; **4**: 6-13.