Immunoinformatics: how in silico methods are re-shaping the investigation of peptide immune specificity

Lawrence Jin Kiat Wee¹,², Shen Jean Lim³, Lisa F.P. Ng¹,³, Joo Chuan Tong²,³

¹Singapore Immunology Network, 8A Biomedical Grove, No. 04-06 Immunos, Biopolis, Singapore 138648, ²Data Mining Department, Institute for Infocomm Research, 1 Fusionopolis Way, No. 21-01 Connexis South Tower, Singapore 138632, ³Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117597

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Immunological databases
4. T-cell epitope prediction
5. B-cell epitope prediction
6. Emerging perspectives
   6.1. Caspases
   6.2. Granzymes
7. Conclusion
8. Acknowledgements
9. References

1. ABSTRACT

In the past decade, information technology has enabled synergistic advances in key domains of immunological research including the development of diagnostics and vaccines. Computational methods of epitope mapping now play instrumental roles in bench experiments, by facilitating the selection of immunogenic targets and the modeling of downstream cellular responses. In this article, we summarize the latest development and application of immune epitope prediction methods and discuss future directions in this field which could enhance our understanding of immune specificity.

2. INTRODUCTION

The immune system serves as the bedrock of the organism’s defense against foreign pathogens. It is made up of two arms – the innate immune response, for immediate non-specific responses to intrusive agents, and the adaptive immune response, for threat-specific responses (1, 2). While the body’s innate immune responses are determined by rapid and instantaneous recognition of a bewildering range of intrusive agents, the adaptive immune responses are characterized by specific memory-dependent assault on the previously identified intrusion. At the heart of the adaptive immune system lies the ability of the organism to
Peptide immunoinformatics

In the T-cell arm of the adaptive immune system, antigenic peptides derived from degradation of cytosolic proteins are bound to the major histocompatibility complex (MHC) class I molecules before being presented to the T-cell receptors on CD8+ cytotoxic T-cells, while peptides derived from degradation of internalized antigens are bound to MHC class II molecules and subsequently recognized by CD4+ helper T-cells. While CD8+ cytotoxic T-cells play a key role in targeted killing of infected or cancerous cells, CD4+ helper T-cells are involved in the initiation and regulation of downstream immuno-signaling responses. Unlike T-cells, B-cells recognize cognate antigens in their native form through binding of the B-cell receptor to epitopes which may be linear or conformational, consisting of distant amino acid sequences brought together spatially upon protein folding. It is believed that about 10% of B-cell epitopes are linear, with the majority being conformational in nature. Together with signaling inputs from other immune cells, B-cells become activated upon binding to the epitope on the antigen and differentiate into mature B-cells which produce and secrete antibodies specific for the antigen.

Due to the inherent combinatorial complexity of the adaptive immune system, as evidenced by the interplay between diverse repertoires of MHC molecules, B-cells, T-cells and antigenic molecules, it is not surprising that experimental studies are complex and could be assisted by data-driven hypothesis generation (3). A primary goal of immunoinformatics is the application of information technology to manage data and model these complex relationships in a high-throughput and systems-wide manner, with the end point of facilitating bench-to-bedside research for vaccine discovery and disease diagnostics (4). In this article, we highlight available immunological databases and relevant resources which are important for investigating immune specificity. Next, we review the latest development and application of computational algorithms for B- and T-cell epitope mapping. Finally, we discuss emerging perspectives on other bioinformatics-based research which could significantly contribute to the investigation of immune specificity.

3. IMMUNOLOGICAL DATABASES AND RESOURCES

High-throughput genome sequencing of the human and other model organism genomes, together with traditional experimental work, have led to a tremendous surge in the availability of biological data. The deluge of data has necessitated the development of specialized immunological data repositories for efficient data storage and retrieval. To date, a total of 27 immunological databases have been archived in the 2010 Nucleic Acids Research Database Collection, ranging from highly specialized, boutique databases to data warehouses integrating data from diverse sources (5). The more prominent resources are highlighted in Table 1.

The Immune Epitope Database and Analysis Resource (IEDB) is a highly integrated web portal containing data related to antibodies and epitopes for humans, non-human primates, rodents, and other animal species (6). It stores over 70,000 entries on epitope sequences related to a diverse range of infectious diseases and allergens. The International ImMunoGeneTics Information System (IMGT) serves as a convenient resource for antibodies, genetic and structural data on the human leukocyte antigen (HLA) molecules, and related proteins of the immune system of human and other vertebrates (7).

Table 1. Biological databases and knowledge portals for immunoinformatics research

<table>
<thead>
<tr>
<th>Resource</th>
<th>Description</th>
<th>URL</th>
<th>Refs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immune Epitope Database (IEDB)</td>
<td>Comprehensive repository of MHC binding peptides, T-cell epitopes and B-cell epitopes data. Also provides variety of MHC-binding and epitope prediction services.</td>
<td><a href="http://www.immuneepitope.org/">http://www.immuneepitope.org/</a></td>
<td>0</td>
</tr>
<tr>
<td>The ImMunoGeneTics International Information System (IMGT)</td>
<td>Highly integrated resource for sequence, structural and genetic information on immune regulators across multiple species. Provide excellent links to relevant resources.</td>
<td><a href="http://imgt.cines.fr/">http://imgt.cines.fr/</a></td>
<td>7</td>
</tr>
<tr>
<td>The Innate Immune Database (IIDD)</td>
<td>Resource for facilitating gene-specific and systems biology oriented research. Enables integrative analysis of individual immune-active genes or the entire genomic locus.</td>
<td><a href="http://db.systemsbiology.net/cgi-bin/GBLUE/U54/IIDDHome.cgi">http://db.systemsbiology.net/cgi-bin/GBLUE/U54/IIDDHome.cgi</a></td>
<td>8</td>
</tr>
<tr>
<td>Immunological Database and Analysis Portal (ImmPort)</td>
<td>Portal for accessing references and experiment data for immunologists. Supports production, analysis, archiving and exchange of scientific data.</td>
<td><a href="https://www.immport.org">https://www.immport.org</a></td>
<td>9</td>
</tr>
<tr>
<td>SYFPETIHH</td>
<td>Database of experimentally verified MHC binding peptides.</td>
<td><a href="http://www.sylpeithi.de/">http://www.sylpeithi.de/</a></td>
<td>0</td>
</tr>
<tr>
<td>MHCBind</td>
<td>Extensive repository of MHC-binding and non-binding peptides.</td>
<td><a href="http://www.imtech.res.in/raghava/mhcbin/">http://www.imtech.res.in/raghava/mhcbin/</a></td>
<td>11</td>
</tr>
<tr>
<td>AntiJen</td>
<td>Database containing quantitative binding data for peptides binding to MHC peptides, T-cell epitopes, TAP, B-cell epitopes and protein-protein interactions.</td>
<td><a href="http://www.imtech.res.in/raghava/antijen/">http://www.imtech.res.in/raghava/antijen/</a></td>
<td>12</td>
</tr>
<tr>
<td>AntigenDB</td>
<td>Comprehensive information about a wide range of experimentally validated antigens cross-linked to epitope data.</td>
<td><a href="http://www.imtech.res.in/raghava/antigendb/">http://www.imtech.res.in/raghava/antigendb/</a></td>
<td>14</td>
</tr>
<tr>
<td>HIV Molecular Immunology Database</td>
<td>HIV-1 cytotoxic and helper T-cell epitopes and antibody binding sites.</td>
<td><a href="http://www.hiv.lanl.gov/content/immunology/">http://www.hiv.lanl.gov/content/immunology/</a></td>
<td>15</td>
</tr>
</tbody>
</table>

Develop and maintain immune specificity to a wide spectrum of immunogenic agents – or antigens, through the T- and B-cells (2).
Table 2. T-cell epitope prediction servers

<table>
<thead>
<tr>
<th>Server</th>
<th>Class</th>
<th>Algorithm(s)</th>
<th>Web Url</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHCPred</td>
<td>I</td>
<td>Multivariate statistical methods</td>
<td><a href="http://www.darriflower.info/mhcppred/">http://www.darriflower.info/mhcppred/</a></td>
<td>15</td>
</tr>
<tr>
<td>NetMHC</td>
<td>I, II</td>
<td>ANN, weight matrix-based</td>
<td><a href="http://www.cbs.dtu.dk/services/NetMHC/">http://www.cbs.dtu.dk/services/NetMHC/</a></td>
<td>15</td>
</tr>
<tr>
<td>PEPITO</td>
<td>I</td>
<td>ANN, matrix-based</td>
<td><a href="http://www.imtech.res.in/raghava/hplpred/neural.html">http://www.imtech.res.in/raghava/hplpred/neural.html</a></td>
<td>20</td>
</tr>
<tr>
<td>SYFPEITHI</td>
<td>I</td>
<td>Matrix-based</td>
<td><a href="http://www.syfpeithi.de/Scripts/MHCServer.dll/EpitopePrediction.html">http://www.syfpeithi.de/Scripts/MHCServer.dll/EpitopePrediction.html</a></td>
<td>21</td>
</tr>
<tr>
<td>ProPred-I</td>
<td>I</td>
<td>Matrix-based</td>
<td><a href="http://www.imtech.res.in/raghava/propred/">http://www.imtech.res.in/raghava/propred/</a></td>
<td>21</td>
</tr>
<tr>
<td>ProPred-II</td>
<td>II</td>
<td>Matrix-based</td>
<td><a href="http://www.imtech.res.in/raghava/propred/">http://www.imtech.res.in/raghava/propred/</a></td>
<td>21</td>
</tr>
<tr>
<td>Rankpep</td>
<td>I</td>
<td>Matrix-based</td>
<td><a href="http://bioinfo2.fori.upf.org/RANKPEP/">http://bioinfo2.fori.upf.org/RANKPEP/</a></td>
<td>21</td>
</tr>
<tr>
<td>NN-align</td>
<td>I</td>
<td>ANN</td>
<td><a href="http://www.cbs.dtu.dk/services/NetMHCII-2.0">http://www.cbs.dtu.dk/services/NetMHCII-2.0</a></td>
<td>21</td>
</tr>
<tr>
<td>SYMHC</td>
<td>I</td>
<td>SVM</td>
<td><a href="http://www-hs.informatik.uni-tuebingen.de/Services/SVMHC">http://www-hs.informatik.uni-tuebingen.de/Services/SVMHC</a></td>
<td>21</td>
</tr>
<tr>
<td>SVRMHC</td>
<td>I</td>
<td>SVM regression</td>
<td><a href="http://svrmhc.biodelead.org/">http://svrmhc.biodelead.org/</a></td>
<td>21</td>
</tr>
<tr>
<td>NetCTL</td>
<td>I</td>
<td>ANN, weight matrix-based</td>
<td><a href="http://www.cbs.dtu.dk/services/NetCTL/">http://www.cbs.dtu.dk/services/NetCTL/</a></td>
<td>21</td>
</tr>
<tr>
<td>MAPP</td>
<td>I</td>
<td>Matrix-based</td>
<td><a href="http://www.mpib-berlin.mpg.de/MAPP/">http://www.mpib-berlin.mpg.de/MAPP/</a></td>
<td>21</td>
</tr>
<tr>
<td>MetaMHC</td>
<td>I</td>
<td>Ensemble method</td>
<td><a href="http://www.biokdd.fudan.edu.cn/Service/MetaMHC.html">http://www.biokdd.fudan.edu.cn/Service/MetaMHC.html</a></td>
<td>21</td>
</tr>
</tbody>
</table>

Another related resource, the Innate Immune Database (IID) provides a useful interface for gene-specific and systems-biology oriented research development, and contains a database of computationally predicted transcription factor binding sites and related genomic features for a set of over 2,000 murine immune genes of interest (8). The Immunological Database and Analysis Portal (ImmPort), provides access to extensive references and experimental data on immunological research, as well as an interface for production, analysis, archival, and exchange of scientific data (9).

The boutique databases such as SYFPEITHI (10) and MHCBN (11) comprise entirely of data on experimentally-derived MHC-binding peptides with more than 7,000 and 20,000 entries respectively. In addition, AntiIen contains over 24,000 entries on experimentally-derived data on MHC peptides, MHC-TCR complexes, T-cell epitopes, transporter associated with antigen processing (TAP) proteins, B-cell epitopes and protein-protein interactions (12). On the other hand, Bcipep focuses on B-cell epitopes binding data and is deposited with over 3,000 entries (13). Unlike the epitope-only databases, AntibGenDB has a compilation of more than 500 antigens culled from literature and other immunological resources (14). These antigens are derived from 44 important pathogenic species where individual antigen entries are annotated with information on sequence, structure and B- and T-cell epitopes. In addition, in silico studies have also benefited from the availability of specialist databases such as the HIV Molecular Immunology Database which has 9,172 records on HIV-1 cytotoxic T-cell epitopes, helper T-cell epitopes and antibody-binding sites, as well as extensive links to HIV-related literature (15).

4. T-CELL EPITOPE PREDICTION

More than 30 T-cell epitope prediction servers have been developed and are available online (16, 17). A number of the more prominent servers are listed in Table 2. As the primary feature in T-cell epitope identification is the requisite binding of the potential peptide to the MHC molecule, majority of the existing predictors are based on identifying such MHC-peptide binding events. The results of prediction range from simple binary outputs to quantitative scores on peptide binding affinity. The pioneering work on MHC-binding prediction is primarily based on experimentally verified epitope motifs (18). For example, in the case of MHC class I binders, it is common to find peptides of length of 8-10 residues with anchoring residues at the N- and C-terminus. Subsequent discovery of the unique contribution of specific peptide residue positions on MHC binding encouraged the development of matrix-based algorithms, such as BIMAS (18) and SYFPEITHI (10). The methods comprised of scoring matrices which quantitatively measure the influence of different amino acids at different residue positions on the overall peptide binding ability to the MHC molecules. However, as matrix-based methods are restricted from accounting the non-linear contributions of residues along the length of the peptide, non-linear algorithms such as Artificial Neural Networks (ANN), Hidden Markov Models (HMM) and Support Vector Machines (SVM) were explored (16, 17). Several prediction servers, including MULTIPRED (19), SVMHC (20) and SVRMHC (21), were developed using these non-linear algorithms and were shown to outperform the matrix methods on independent testing (16, 17). More recently, servers such as NetMHC2 (22) have shown that integration of two or more prediction methods either by averaging over the predictions made or by feeding the prediction outputs from one method to another, could lead to better overall prediction performance.

While much progress has been made in designing accurate MHC class I predictors, development of MHC class II predictors have been complicated by several factors (23). Structurally, the open binding cleft of MHC class II molecules allows for greater degeneracy in the length of the binding peptides – and consequently a much more varied T-cell epitope repertoire is observed. In addition, the MHC class II binding motifs have relatively weak and often degenerate sequence signals. To date, most of the methods for MHC class II binding predictions have been trained and evaluated on very limited datasets covering only a single or a few different MHC class II alleles. Hence, there are...
Peptide immunoinformatics

Table 3. B-cell epitope prediction servers

<table>
<thead>
<tr>
<th>Server</th>
<th>Algorithm</th>
<th>Web Url</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>RecPred</td>
<td>FPS</td>
<td><a href="http://www.imtech.res.in/raghava/bcppred/">http://www.imtech.res.in/raghava/bcppred/</a></td>
<td>33</td>
</tr>
<tr>
<td>BepiPred</td>
<td>FPS, HMM</td>
<td><a href="http://www.cbs.dtu.dk/services/BepiPred">http://www.cbs.dtu.dk/services/BepiPred</a></td>
<td>37</td>
</tr>
<tr>
<td>ABCPred</td>
<td>ANN</td>
<td><a href="http://www.imtech.res.in/raghava/abcppred/">http://www.imtech.res.in/raghava/abcppred/</a></td>
<td>42</td>
</tr>
<tr>
<td>BCPredS</td>
<td>SVM, String kernels</td>
<td><a href="http://ailab.cs.iastate.edu/bcppreds/">http://ailab.cs.iastate.edu/bcppreds/</a></td>
<td>39</td>
</tr>
<tr>
<td>COBEpro</td>
<td>SVM, Sequence similarity</td>
<td><a href="http://scratch.proteomics.ics.uci.edu/">http://scratch.proteomics.ics.uci.edu/</a></td>
<td>38</td>
</tr>
<tr>
<td>Discotope</td>
<td>Structural-based</td>
<td><a href="http://tools.immuneepitope.org/tools/discotope/discotope.do">http://tools.immuneepitope.org/tools/discotope/discotope.do</a></td>
<td>44</td>
</tr>
<tr>
<td>Ellipro</td>
<td>Structural-based</td>
<td><a href="http://tools.immuneepitope.org/tools/Ellipro/ellib_input">http://tools.immuneepitope.org/tools/Ellipro/ellib_input</a></td>
<td>45</td>
</tr>
<tr>
<td>Epito (Pepto)</td>
<td>Amino-acid propensities, structural</td>
<td><a href="http://www.igb.uci.edu/">http://www.igb.uci.edu/</a></td>
<td>46</td>
</tr>
<tr>
<td>Epitopia</td>
<td>Naive Bayes classifier</td>
<td><a href="http://epitopia.tau.ac.il/">http://epitopia.tau.ac.il/</a></td>
<td>0,41</td>
</tr>
</tbody>
</table>

Besides modeling MHC-binding events, significant work have been done on developing systems that model and integrate the events upstream of MHC peptide presentation. Notably, a number of servers are available for predicting antigen processing and peptide transport through the MHC class I presentation pathway. FRAGPREDICT (26), PAPoC (27), NetChop (28) and Peleavag (29) are dedicated methods for proteasomal cleavage prediction. FRAGPREDICT utilizes a motif-based algorithm and an experimentally-defined kinetic model for proteasomal cleavage prediction. On the other hand, PAPoC adopts a stochastic hill-climbing algorithm while NetChop and Peleavag are developed using ANN and SVM algorithms respectively. Interestingly, servers such as MAPPP (30) and NetCTL (31) have incorporated multiple prediction services for integrated modeling of the various molecular events leading up to MHC-peptide interaction. For the MAPPP server, potential MHC class I binding peptides are predicted for proteasomal cleavage using either the FRAGPREDICT or PAPoC algorithms while MHC-binding prediction are made using the BIMAS or SYFPEITHI methods. For NetCTL, MHC-peptide binding is predicted using ANN and weight matrices while proteasomal cleavage prediction is based on the NetChop algorithm. In addition, NetCTL offers a predicted output on TAP-mediated transport efficiency of the query peptide using an experimentally-derived weight matrix.

5. B-CELL EPITOPE PREDICTION

While T-cell epitope prediction have attained significant performance and could be suitably applied for preliminary epitope mapping studies, progress in predicting B-cell epitopes has been considerably slower. Due to the highly varied nature of epitope binding to the B-cell receptor, it is expected that the accurate modeling of the epitope and receptor interaction would be significantly more complex. Nonetheless, a wide range of B-cell prediction methods have been developed (Table 3), with much of the current research being devoted to the prediction of linear B-cell epitopes. Early efforts in this field were primarily based on the use of amino acid propensity scales to identify amino acid residues that are most commonly found in B-cell epitope sequences. Computational methods that implement such scales include PREDITOP (32), PEOPLE (33), BEPITOPE (34) and BcPred (35). However, the effectiveness of amino acid propensity scales in detecting B-cell epitopes remains a subject of debate. An extensive assessment by Blythe and Flower (36) on 484 propensity scales concluded that even the best set of scales and parameters performed only marginally better than random and cannot be used to predict epitope location reliably. This has led to the development of more sophisticated methods, some of which incorporate machine learning techniques, to address the growing need for reliable prediction. Specific examples include BepiPred, which combines HMM with propensity scale methods (37); COBEpro, which employs SVM and fragment predictions to compute an epitopic propensity score for each residue (38); BCpredS, which utilized SVM together with string kernels for predicting linear peptides of 12-20 amino acids in length (39); Epitopia, which uses a naïve Bayes classifier to predict immunogenic regions on either a protein 3D structure or linear sequence (40, 41); and ABCPred, which employs a recurrent neural network to predict continuous B-cell epitopes on the antigen (42). Decision tree based models have also been reported and used for analysis of protective continuous epitopes (43).

Much progress has been made in the development of discontinuous or conformational epitope prediction algorithms, most of which harness structure-based technologies to analyze the protein’s globular surface. One good example is the Discotope server, which was developed using amino acid statistics, spatial information, and surface accessibility on a set of experimentally resolved discontinuous epitope structures (44). Others include Ellipro, which implements Thornton’s method, a residue clustering algorithm and homology modeling algorithm for epitope screening (45), and PEPITO (46), which utilizes a combination of amino-acid propensity scores and half sphere exposure values at multiple distances for prediction. With the rapid growth of experimental 3-D structures in the Protein Data Bank (PDB) (47), it is expected that more structure-based methods will emerge for the development of linear and conformational B-cell epitope prediction algorithms.

6. EMERGING PERSPECTIVES

While most immunoinformatics research have been centered on the traditional aspects of immunology
Peptide immunoinformatics

such as MHC-peptide binding and epitope prediction, a number of recent experimental studies have highlighted the potential of integrating bioinformatics applications from other domains to refine the investigation of peptide immune specificity. In the following sections, we summarize these findings and review the in silico methods available.

6.1. Caspases

Caspases belong to a unique class of cysteine proteases which function as critical effectors of apoptosis, inflammation and other important cellular processes such as cell proliferation and cell differentiation (48, 49). Caspases cleave substrates at specific tetrapeptide sites with a highly conserved aspartate (D) at the P1 position (50). To date, more than 300 different caspase substrates have been experimentally defined (51). These substrates belong to a myriad of functional classes such as cell cycle regulators, DNA-binding proteins, cell surface receptors and viral proteins. In a recent study by Rawson et al. (52), the first evidence of caspase involvement in immunopathology was noted when they found that the proteome of apoptotic T-cells included fragments of cellular proteins generated by caspases and that a high proportion of distinct T-cell epitopes in these fragments were recognized by CD8+ cytotoxic T-cells during HIV infection. The frequencies of CD8+ cytotoxic T-cells that are specific for apoptosis-dependent epitopes correlate with the frequency of circulating apoptotic CD4+ helper T-cells in HIV-1-infected individuals. It was further suggested that caspase-dependent cleavage of proteins associated with apoptotic cells has a key role in the induction of self-reactive CD8+ cytotoxic T-cell responses, as the caspase-cleaved fragments are efficiently targeted to the processing machinery and are cross-presented by dendritic cells. In addition, Lopez et al. reported that caspases were involved in processing and presentation of a short vaccinia virus-encoded antigen (53). By cleaving at non-canonical sites, at least two caspases were found to generate antigenic peptides recognized by the T-cells. As the cleavage sites and peptide products were partially overlapping but different to those produced by proteasomes in vitro, it was suggested that caspase-mediated cleavage might be an alternative mode of antigen processing.

A number of online servers are available for caspase cleavage site prediction (Table 4). PeptideCutter is a general proteolytic cleavage prediction server which has in-built modules for a number of different caspases based on expertly curated cleavage motifs (54). Lohmuller et al. developed the peptidase substrate prediction tool (PEPS) based on position-specific scoring matrices (PSSM) for cathepsin B, cathepsin L and caspase-3 substrates (55). Garay-Malpartida et al. (56) developed the CasPredictor software based on a similar PSSM and the GraBCas software by Backes et al. (57) advanced the earlier PSSM-based methods by training on an updated set of caspase cleavage specificities. More recently, machine-learning algorithms have been implemented for caspase cleavage prediction and were shown to perform better than the earlier motif- and matrix-based methods. Wee et al., developed a SVM-based method utilizing various sequence lengths for prediction and incorporated secondary structure and solvent accessibility features (58-60). CasCleave, another SVM-based prediction server, was developed using Bayes Feature Extraction for feature representation (61). More recently, Pripper utilized a variety of machine learning algorithms, including SVM, J48 and random forest, for caspase cleavage prediction of whole proteomes (62).

6.2. Granzymes

Granule enzymes, or granzymes, belong to a unique class of serine proteases which play critical roles in the immune response through the killing of virus-infected or tumor cells (63). Granzymes are released into the cytoplasm of the target cells through endocytosis of cytolytic granules released by cytotoxic T-cells and natural killer (NK) cells. Once released into the target cells, granzymes go on to cleave specific cellular proteins which activate multiple signaling pathways leading to apoptotic cell death. The most well studied granzyme, granzyme B, is known to recognize and cleave proteins at specific tetrapeptide motifs with Asp (D) residue at the P1 position.

Although granzyme-induced apoptotic cell death has long been considered the de facto mechanism for killing virus-infected cells, accumulating evidence suggest that granzymes also mediate antiviral effects through distinctive non-apoptotic pathways. Andrade et al. reported that the adenovirus type 5 DNA-binding protein and the 100K protein are cleaved by granzyme B and granzyme H (64). These proteins are essential for adenovirus replication and cleavage by the granzymes was shown to inactivate these proteins and inhibit viral replication. It was also reported that elevated concentrations of circulating granzymes were found in various inflammatory processes and that granzymes could mediate cleavage of extracellular substrates (reviewed in ref. 65). Together, these findings suggest that granzymes mediate a board range of functions relevant to antiviral activities and tumor rejection, as well as the pathogenesis of chronic inflammatory diseases. As reviewed in Darrah et al. (66), granzyme-mediated cleavage was found to modify the structure of autoantigens during cytolytic granule-mediated cell death and could be instrumental in driving the progression of systematic autoimmune diseases such as systemic lupus erythematosus and rheumatoid arthritis. Granzyme B cleavage sites were found to co-localize with autoimmune epitopes and cleavage of cellular proteins has been shown to create or destroy the autoimmune epitopes. There is much evidence to suggest that elucidation of granzyme targets and their cleavage products will have profound impact on the
Peptide immunoinformatics

understanding of the role of granzymes in immune responses. However, unlike caspase cleavage prediction, there are limited resources for predicting granzyme cleavage sites. Both GraBCas and PeptideCutter are available for granzyme B cleavage site prediction, in addition to caspase cleavage prediction. GraBCas uses a position-specific scoring matrix model derived from quantitative measures of cleavage specificities of granzyme B (57), while PeptideCutter employs expertly curated cleavage motifs for prediction (54).

7. CONCLUSION

With the ever-growing availability and deposition of biological data, it clear that information technology has a critical role to play in driving various aspects of immunology research. It is expected that epitope discovery will continue to be the mainstay in defining immune specificity and that these efforts will be complemented by sophisticated in silico methods, which are increasingly being integrated into platforms for system-wide analyses. Synergistic interactions between experimental and computational research will augment practical efforts in applied immunological research such as vaccine and diagnostics development.

8. ACKNOWLEDGEMENTS

This work was supported by a joint council research grant from the Joint Council Office (JCO) of A*STAR Singapore.

9. REFERENCES

18. KC Parker, MA Bednarek, JE Coligan: Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J Immunol 152, 163-175 (1994)
Peptide immunoinformatics


Peptide immunoinformatics


68. P Guan, CK Hattotuwagama, IA Doytchinova, DR Flower: MHCPred 2.0: an updated quantitative T-cell epitope prediction server. *Appl Bioinformatics* 5, 55-61 (2006)


Peptide immunoinformatics


**Abbreviations**: MHC: major histocompatibility complex; HLA: human leukocyte antigen; NK: natural killer; TAP: transporter associated with antigen processing; ANN: artificial neural network; HMM: hidden Markov model; SVM: support vector machine; PPS: physicochemical propensity scale

**Key Words** Bioinformatics, Immunoinformatics, Immunological Databases, T-cell epitope prediction, B-cell Epitope Prediction, Ig, immunoglobulin; TR: T-cell receptor, Review

**Send correspondence to**: Joo Chuan Tong, Data Mining Department, Institute for Infocomm Research, 1 Fusionopolis Way, No. 21-01 Connexis South Tower, Singapore 138632, Tel: 65-6408-2156, Fax: 65-6776-1378, E-mail: victor@bic.nus.edu.sg

http://www.bioscience.org/current/volE4.htm