

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

**Lawrence Jin Kiat Wee<sup>1,2</sup>, Shen Jean Lim<sup>3</sup>, Lisa F.P. Ng<sup>1,3</sup>, Joo Chuan Tong<sup>2,3</sup>**

<sup>1</sup>Singapore Immunology Network, 8A Biomedical Grove, No. 04-06 Immunos, Biopolis, Singapore 138648, <sup>2</sup>Data Mining Department, Institute for Infocomm Research, 1 Fusionopolis Way, No. 21-01 Connexis South Tower, Singapore 138632, <sup>3</sup>Department 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

**Table 1.** Biological databases and knowledge portals for immunoinformatics research

Resource	Description	URL	Refs.
Immune Epitope Database (IEDB)	Comprehensive repository of MHC binding peptides, T-cell epitopes and B-cell epitopes data. Also provides a variety of MHC-binding and epitope prediction services.	<a href="http://www.immuneepitope.org/">http://www.immuneepitope.org/</a>	6
The International ImMunoGeneTics system (IMGT)	Highly integrated resource for sequence, structural and genetic information on immune regulators across multiple species. Provide excellent links to relevant resources.	<a href="http://imgt.cines.fr/">http://imgt.cines.fr/</a>	7
The Innate Immune Database (IIDB)	Resource for facilitating gene-specific and systems biology oriented research. Enables integrative analysis of individual immune-active genes or the entire genomic locus.	<a href="http://db.systemsbio.net/cgi-bin/GLUE/U54/IIDBHome.cgi">http://db.systemsbio.net/cgi-bin/GLUE/U54/IIDBHome.cgi</a>	8
Immunological Database and Analysis Portal (ImmPort)	Portal for accessing references and experiment data for immunologists. Supports production, analysis, archiving, and exchange of scientific data.	<a href="https://www.immport.org">https://www.immport.org</a>	9
SYFPEITHI	Database of experimentally verified MHC binding peptides.	<a href="http://www.syfpeithi.de/">http://www.syfpeithi.de/</a>	10
MHCBN	Extensive repository of MHC-binding and non-binding peptides.	<a href="http://www.imtech.res.in/raghava/mhcbn/">http://www.imtech.res.in/raghava/mhcbn/</a>	11
AntiJen	Database containing quantitative binding data for peptides binding to MHC peptides, T-cell epitopes, TAP, B-cell epitopes and protein-protein interactions.	<a href="http://www.darrenflower.info/antijen/">http://www.darrenflower.info/antijen/</a>	12
Bcipep	Extensive repository of B-cell epitopes.	<a href="http://www.imtech.res.in/raghava/bcipep/">http://www.imtech.res.in/raghava/bcipep/</a>	13
AntigenDB	Comprehensive information about a wide range of experimentally-validated antigens cross-linked to epitope data.	<a href="http://www.imtech.res.in/raghava/antigendb/">http://www.imtech.res.in/raghava/antigendb/</a>	14
HIV Molecular Immunology Database	HIV-1 cytotoxic and helper T-cell epitopes and antibody binding sites.	<a href="http://www.hiv.lanl.gov/content/immunology/">http://www.hiv.lanl.gov/content/immunology/</a>	15

develop and maintain immune specificity to a wide spectrum of immunogenic agents – or antigens, through the T- and B-cells (2).

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<sup>+</sup> cytotoxic T-cells, while peptides derived from degradation of internalized antigens are bound to MHC class II molecules and subsequently recognized by CD4<sup>+</sup> helper T-cells. While CD8<sup>+</sup> cytotoxic T-cells play a key role in targeted killing of infected or cancerous cells, CD4<sup>+</sup> 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).

IMGT currently (November 2010) contains six databases: (i) IMGT/LIGM-DB with 150,027 immunoglobulin (IG) and T-cell receptor (TCR) sequences from 261 species; (ii) IMGT/MHC-DB with sequences of 2,292 HLA class I alleles, 1,012 HLA class II alleles, and 106 non-HLA alleles; (iii) IMGT/GENE-DB with 2,702

**Table 2.** T-cell epitope prediction servers

Server	Class	Algorithm(s)	Web Url	Refs.
BIMAS		Matrix-based	<a href="http://www.bimas.cit.nih.gov/molbio/hla_bind/">http://www.bimas.cit.nih.gov/molbio/hla_bind/</a>	8
IEDB	II	Matrix-based, ANN	<a href="http://tools.immuneepitope.org/main/html/tcell_tools.html">http://tools.immuneepitope.org/main/html/tcell_tools.html</a>	9
MHC-I		Structural-based with protein threading	<a href="http://atom.research.microsoft.com/hlabinding/hlabinding.aspx">http://atom.research.microsoft.com/hlabinding/hlabinding.aspx</a>	7
MHCPred	II	Multivariate statistical methods	<a href="http://www.darrenflower.info/mhcpred/">http://www.darrenflower.info/mhcpred/</a>	8
NetMHC		ANN, weight matrix-based	<a href="http://www.cbs.dtu.dk/services/NetMHC/">http://www.cbs.dtu.dk/services/NetMHC/</a>	12
nHLAPred		ANN, matrix-based	<a href="http://www.imtech.res.in/raghava/nhlapred/neural.html">http://www.imtech.res.in/raghava/nhlapred/neural.html</a>	10
SYFPEITHI	II	Matrix-based	<a href="http://www.syfpeithi.de/Scripts/MHCServer.dll/EpitopePrediction.htm">http://www.syfpeithi.de/Scripts/MHCServer.dll/EpitopePrediction.htm</a>	10
ProPred-I		Matrix-based	<a href="http://www.imtech.res.in/raghava/propred1/">http://www.imtech.res.in/raghava/propred1/</a>	11
ProPred-II	I	Matrix-based	<a href="http://www.imtech.res.in/raghava/propred/">http://www.imtech.res.in/raghava/propred/</a>	12
Rankpep	II	Matrix-based	<a href="http://bio.dfci.harvard.edu/RANKPEP/">http://bio.dfci.harvard.edu/RANKPEP/</a>	13
NN-align	I	ANN	<a href="http://www.cbs.dtu.dk/services/NetMHCII-2.0">http://www.cbs.dtu.dk/services/NetMHCII-2.0</a>	14
SVMHC	II	SVM	<a href="http://www.bs.informatik.uni-tuebingen.de/Services/SVMHC">http://www.bs.informatik.uni-tuebingen.de/Services/SVMHC</a>	10
SVRMHC	II	SVM regression	<a href="http://svrmhc.biocloud.org/">http://svrmhc.biocloud.org/</a>	11
NetCTL		ANN, weight matrix-based	<a href="http://www.cbs.dtu.dk/services/NetCTL/">http://www.cbs.dtu.dk/services/NetCTL/</a>	11
MAPPP		Matrix-based	<a href="http://www.mpiib-berlin.mpg.de/MAPPP/">http://www.mpiib-berlin.mpg.de/MAPPP/</a>	10
MetaMHC	II	Ensemble method	<a href="http://www.biokdd.fudan.edu.cn/Service/MetaMHC.html">http://www.biokdd.fudan.edu.cn/Service/MetaMHC.html</a>	14

genes and 3,761 alleles of human, mouse, rat and rabbit IG and TCR genes; (iv) IMGT/PRIMER-DB with 1,864 primer records of IG and TCR from 11 species; (v) IMGT/3Dstructure-DB with 2,367 records of IG, TCR, and MHC proteins with known 3D structures; and (vi) IMGT/mAb-DB with 343 entries of monoclonal antibodies and fusion proteins for immune applications.

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, AntiJen 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 epitopes-only databases, AntigenDB 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 NetMHC (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

**Table 3.** B-cell epitope prediction servers

Server	Algorithm	Web Url	Refs.
BcePred	PPS	<a href="http://www.imtech.res.in/raghava/bcepred/">http://www.imtech.res.in/raghava/bcepred/</a>	35
BepiPred	PPS, HMM	<a href="http://www.cbs.dtu.dk/services/BepiPred">http://www.cbs.dtu.dk/services/BepiPred</a>	37
ABCPred	ANN	<a href="http://www.imtech.res.in/raghava/abcpred/">http://www.imtech.res.in/raghava/abcpred/</a>	42
BCPredS	SVM, String kernels	<a href="http://ailab.cs.iastate.edu/bcpreds/">http://ailab.cs.iastate.edu/bcpreds/</a>	39
COBEpro	SVM, Sequence similarity	<a href="http://scratch.proteomics.ics.uci.edu/">http://scratch.proteomics.ics.uci.edu/</a>	38
Discotope	Structural-based	<a href="http://tools.immuneepitope.org/tools/discotope/discotope.do">http://tools.immuneepitope.org/tools/discotope/discotope.do</a>	44
ElliPro	Structural-based	<a href="http://tools.immuneepitope.org/tools/ElliPro/iedb_input">http://tools.immuneepitope.org/tools/ElliPro/iedb_input</a>	45
Bepro (Pepito)	Amino-acid propensities, structural	<a href="http://www.igb.uci.edu/">http://www.igb.uci.edu/</a>	46
Epitopia	Naïve Bayes classifier	<a href="http://epitopia.tau.ac.il/">http://epitopia.tau.ac.il/</a>	40,41

correspondingly less available servers for MHC class II binding prediction and a much more limited adoption of these methods for MHC class II epitope discovery (23). However, it is encouraging to note that ensemble-based methods – such as MetaMHC - were found to perform better when compared to the use of individual algorithms on their own (24). The MetaMHC server aggregates prediction outputs from distinctive, standalone MHC-binding prediction algorithms and computes the prediction outcome. As these results are in agreement with other ensemble-based methods in related computational domains (25), it is expected that more algorithmic work would be carried from this perspective.

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), PAProC (27), NetChop (28) and Pcleavage (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, PAProC adopts a stochastic hill-climbing algorithm while NetChop and Pcleavage 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 PAProC 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 of 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 BcePred (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

**Table 4.** Caspases and granzymes cleavage sites prediction servers

Server	Algorithm	Web Url	Refs.
PeptideCutter	Motif-based	<a href="http://www.expasy.ch/tools/peptidecutter/">http://www.expasy.ch/tools/peptidecutter/</a>	54
CASVM	SVM	<a href="http://www.casbase.org/casvm/index.html">http://www.casbase.org/casvm/index.html</a>	58-60
CasCleave	SVM, Bayes Feature Extraction	<a href="http://sunflower.kuicr.kyoto-u.ac.jp/~sjn/CasCleave/">http://sunflower.kuicr.kyoto-u.ac.jp/~sjn/CasCleave/</a>	61
Pripper	SVM, Random Forest, J48	<a href="http://users.utu.fi/mijopi/Pripper">http://users.utu.fi/mijopi/Pripper</a>	62

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 P<sub>1</sub> 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<sup>+</sup> cytotoxic T-cells during HIV infection. The frequencies of CD8<sup>+</sup> cytotoxic T-cells that are specific for apoptosis-dependent epitopes correlate with the frequency of circulating apoptotic CD4<sup>+</sup> 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<sup>+</sup> 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 P<sub>1</sub> 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

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

1. T Saitoh, S Akira: Regulation of innate immune responses by autophagy-related proteins. *J Cell Biol* 189, 925-935 (2010)
2. S Paust, B Senman, UH von Andrian: Adaptive immune responses mediated by natural killer cells. *Immunol Rev* 235, 286-296 (2010)
3. Q Yan: Immunoinformatics and systems biology methods for personalized medicine. *Methods Mol Biol* 662, 203-220 (2010)
4. S Vivona, JL Gardy, S Ramachandran, FS Brinkman, GP Raghava, DR Flower, F Filippini: Computer-aided biotechnology: from immuno-informatics to reverse vaccinology. *Trends Biotechnol* 26, 190-200 (2008)
5. GR Cochrane, MY Galperin: The 2010 Nucleic Acids Research Database Issue and online Database Collection: a community of data resources. *Nucleic Acids Res* 38, D1-4 (2010)
6. R Vita, L Zarebski, JA Greenbaum, H Emami, I Hoof, N Salimi, R Damle, A Sette, B Peters: The immune epitope database 2.0. *Nucleic Acids Res* 38, D854-862 (2010)
7. MP Lefranc, V Giudicelli, L Regnier, P Duroux: IMGT, a system and an ontolog that bridge biological and computational spheres in bioinformatics. *Brief Bioinform* 9, 263-275 (2008)
8. M Korb, AG Rust, V Thorsson, C Battail, B Li, D Hwang, KA Kennedy, JC Roach, CM Rosenberger, M Gilchrist, D Zak, C Johnson, B Marzolf, A Aderem, I Shmulevich, H Bolouri: The Innate Immune Database (IIDB). *BMC Immunol* 9, 7 (2008)
9. Y Megan Kong, C Dahlke, Q Xiang, Y Qian, D Karp, RH Scheuermann: Toward an ontology-based framework for clinical research databases. *J Biomed Inform*, in press.
10. MM Schuler, MD Nastke, S Stevanovik: SYFPEITHI: database for searching and T-cell epitope prediction. *Methods Mol Biol* 409, 75-93 (2007)
11. S Lata, M Bhasin, GP Raghava: MHCBN 4.0: A database of MHC/TAP binding peptides and T-cell epitopes. *BMC Res Notes* 2, 61 (2009)
12. CP Toseland, DJ Clayton, H McSparron, SL Hemsley, MJ Blythe, K Paine, IA Doytchinova, P Guan, CK Hattotuagama, DR Flower: AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data. *Immunome Res* 1, 4 (2005)
13. S Saha, M Bhasin, GP Raghava: Bcipep: a database of B-cell epitopes. *BMC Genomics* 6, 79 (2005)
14. HR Ansari, DR Flower, GP Raghava: AntigenDB: an immunoinformatics database of pathogen antigens. *Nucleic Acids Res* 38, D847-853 (2010)
15. HIV Molecular Immunology 2006/2007, Eds: BTM Korber, C Brander, BF Haynes, R Koup, JP Moore, BD Walker, DI Watkins, *Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico. LA-UR 07-4752* (2006/2007)
16. HH Lin, S Ray, S Tongchusak, EL Reinherz, V Brusic: Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. *BMC Immunol* 9, 8 (2008)
17. HH Lin, GL Zhang, S Tongchusak, EL Reinherz, V Brusic: Evaluation of MHC-II peptide binding prediction servers: applications for vaccine research. *BMC Bioinformatics* 9, S22 (2008)
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)
19. GL Zhang, AM Khan, KN Srinivasan, JT August, V Brusic: MULTIPRED: a computational system for prediction of promiscuous HLA binding peptides. *Nucleic Acids Res* 33, W172-179 (2005)

20. P Dönnes, O Kohlbacher: SVMHC: a server for prediction of MHC-binding peptides. *Nucleic Acids Res* 34, W194-197 (2006)
21. J Wan, W Liu, Q Xu, Y Ren, DR Flower, T Li: SVRMHC prediction server for MHC-binding peptides. *BMC Bioinformatics* 7, 463 (2006)
22. C Lundegaard, K Lamberth, M Harndahl, S Buus, O Lund, M Nielsen: NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Res* 36, W509-512 (2008)
23. C Lundegaard, O Lund, C Kesmir, S Brunak, M Nielsen: Modeling the adaptive immune system: predictions and simulations. *Bioinformatics* 23, 3265-3275 (2007)
24. X Hu, W Zhou, K Udaka, H Mamitsuka, S Zhu: MetaMHC: a meta approach to predict peptides binding to MHC molecules. *Nucleic Acids Res* 38, W474-479 (2010)
25. R Polikar: Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine* 6, 21-45 (2006)
26. HG Holzhütter, C Frömmel, PM Kloetzel: A theoretical approach towards the identification of cleavage-determining amino acid motifs of the 20 S proteasome. *J Mol Biol* 286, 1251-1265 (1999)
27. AK Nussbaum, C Kuttler, KP Haderl, HG Rammensee, H Schild: PAMProC: a prediction algorithm for proteasomal cleavages available on the WWW. *Immunogenetics* 53, 87-94 (2001)
28. C Kesmir, AK Nussbaum, H Schild, V Detours, S Brunak: Prediction of proteasome cleavage motifs by neural networks. *Protein Eng* 15, 287-296 (2002)
29. M Bhasin, GP Raghava: Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. *Nucleic Acids Res* 33, W202-207 (2005)
30. J Hakenberg, AK Nussbaum, H Schild, HG Rammensee, C Kuttler, HG Holzhütter, PM Kloetzel, SH Kaufmann, HJ Mollenkopf: MAPPP: MHC class I antigenic peptide processing prediction. *Appl Bioinformatics* 2, 155-158 (2003)
31. T Stranzl, MV Larsen, C Lundegaard, M Nielsen: NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics* 62, 357-368 (2010)
32. JL Pellequer, E Westhof: PREDITOP: a program for antigenicity prediction. *J Mol Graph* 11, 204-210 (1993)
33. AJ Alix: Predictive estimation of protein linear epitopes by using the program PEOPLE. *Vaccine* 18, 311-314 (1999)
34. M Odorico, JL Pellequer: BEPITOPE: predicting the location of continuous epitopes and patterns in proteins. *J Mol Recognit* 16, 20-22 (2003)
35. S Saha, GP Raghava: BcePred: Prediction of continuous B-Cell epitopes in antigenic sequences using physico-chemical properties. In: ICARIS. Eds: G Nicosia, V Cutello, PJ Bentley, J Timis, Springer (2004)
36. MJ Blythe, DR Flower: Benchmarking B cell epitope prediction underperformance of existing methods. *Protein Sci* 14, 246-248 (2005)
37. JE Larsen, O Lund, M Nielsen: Improved method for predicting linear B-cell epitopes. *Immunome Res* 2, 2 (2006)
38. MJ Sweredoski, P Baldi: COBEpro: a novel system for predicting continuous B-cell epitopes. *Protein Eng Des Sel* 22, 113-120 (2009)
39. Y El-Manzalawy, D Dobbs, V Honavar: Predicting linear B-cell epitopes using string kernels. *J Mol Recognit* 4, 243-255 (2008)
40. ND Rubinstein, I Mayrose, T Pupko: A machine-learning approach for predicting B-cell epitopes. *Mol Immunol* 46, 840-847 (2009)
41. ND Rubinstein, I Mayrose, E Martz, T Pupko: Epitopia: a web-server for predicting B-cell epitopes. *BMC Bioinformatics* 10, 287 (2009)
42. S Saha, GP Raghava: Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 65, 40-48 (2006)
43. J Söller, B Mayer: Machine learning approaches for prediction of linear B-cell epitopes on proteins. *J Mol Recognit* 19, 200-208 (2006)
44. P Haste Andersen, M Nielsen, O Lund: Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 15, 2558-2567 (2006)
45. J Ponomarenko, HH Bui, W Li, N Fusseder, PE Bourne, A Sette, B Peters: ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* 9, 514 (2008)
46. MJ Sweredoski, P Baldi: PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics* 24, 1459-1460 (2008)
47. FC Bernstein, TF Koetzle, GJ Williams, EF Jr Meyer, MD Brice, JR Rodgers, O Kennard, T Shimanouchi, M Tasumi: The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol* 112, 535-542 (1977)
48. M Los, C Stroh, RU Jänicke, IH Engels, K Schulze-Osthoff: Caspases: more than just killers? *Trends Immunol* 22, 31-34 (2001)

49. S Launay, O Hermine, M Fontenay, G Kroemer, E Solary, C Garrido: Vital functions for lethal caspases. *Oncogene* 24, 5137-5148 (2005)
50. WC Earnshaw, LM Martins, SH Kaufmann: Mammalian caspases: structure, activation, substrates, and functions during apoptosis. *Annu Rev Biochem* 68, 383-424 (1999)
51. U Fischer, RU Jänicke, K Schulze-Osthoff: Many cuts to ruin: a comprehensive update of caspase substrates. *Cell Death Differ* 10, 76-100 (2003)
52. PM Rawson, C Molette, M Videtta, L Altieri, D Franceschini, T Donato, L Finocchi, A Propato, M Paroli, F Meloni, CM Mastroianni, G d'Ettore, J Sidney, A Sette, V Barnaba: Cross-presentation of caspase-cleaved apoptotic self antigens in HIV infection. *Nat Med* 13, 1431-1439 (2007)
53. D López, M García-Calvo, GL Smith, M Del Val: Caspases in virus-infected cells contribute to recognition by CD8+ T lymphocytes. *J Immunol* 184, 5193-5199 (2010)
54. E Gasteiger, C Hoogland, A Gattiker, S Duvaud, MR Wilkins, RD Appel, A Bairoch: Protein identification and analysis tools on the ExPASy server. In: The proteomics protocols handbook. Ed: Walker JM, Humana Press (2005)
55. T Lohmüller, D Wenzler, S Hagemann, W Kiess, C Peters, T Dandekar, T Reinheckel: Toward computer-based cleavage site prediction of cysteine endopeptidases. *Biol Chem* 384, 899-909 (2003)
56. HM Garay-Malpartida, JM Occhiucci, J Alves, JE Belizário: CaSPredictor: a new computer-based tool for caspase substrate prediction. *Bioinformatics* 21, i169-176 (2005)
57. C Backes, J Kuentzer, HP Lenhof, N Comtesse, E Meese: GraBCas: a bioinformatics tool for score-based prediction of Caspase- and Granzyme B-cleavage sites in protein sequences. *Nucleic Acids Res* 33, W208-213 (2005)
58. LJ Wee, TW Tan, S Ranganathan: SVM-based prediction of caspase substrate cleavage sites. *BMC Bioinformatics* 7, S14 (2006)
59. LJ Wee, TW Tan, S Ranganathan: CASVM: web server for SVM-based prediction of caspase substrates cleavage sites. *Bioinformatics* 23, 3241-3243 (2007)
60. LJ Wee, JC Tong, TW Tan, S Ranganathan: A multi-factor model for caspase degradome prediction. *BMC Genomics* 10, S6 (2009)
61. J Song, H Tan, H Shen, K Mahmood, SE Boyd, GI Webb, T Akutsu, JC Whisstock: Cascleave: towards more accurate prediction of caspase substrate cleavage sites. *Bioinformatics* 26, 752-760 (2010)
62. M Piippo, N Lietzén, OS Nevalainen, J Salmi, TA Nyman: Prippe: prediction of caspase cleavage sites from whole proteomes. *BMC Bioinformatics* 11, 320 (2010)
63. KM Heutinck, IJ ten Berge, CE Hack, J Hamann, AT Rowshani: Serine proteases of the human immune system in health and disease. *Mol Immunol* 47, 1943-1955 (2010)
64. F Andrade, E Fellows, DE Jenne, A Rosen, CS Young: Granzyme H destroys the function of critical adenoviral proteins required for viral DNA replication and granzyme B inhibition. *EMBO J* 26, 2148-2157 (2007)
65. V Romero, F Andrade: Non-apoptotic functions of granzymes. *Tissue Antigens* 71, 409-416 (2008)
66. E Darrah, A Rosen: Granzyme B cleavage of autoantigens in autoimmunity. *Cell Death Differ* 17, 624-632 (2010)
67. N Jojic, M Reyes-Gomez, D Heckerman, C Kadie, O Schueler-Furman: Learning MHC I – peptide binding. *Bioinformatics* 22, e227-235 (2006)
68. P Guan, CK Hattotuwigama, IA Doytchinova, DR Flower: MHCpred 2.0: an updated quantitative T-cell epitope prediction server. *Appl Bioinformatics* 5, 55-61 (2006)
69. GL Zhang, AM Khan, KN Srinivasan, JT August, V Brusic: MULTIPRED: a computational system for prediction of promiscuous HLA binding peptides. *Nucleic Acids Res* W172-179 (2005)
70. M Bhasin, GP Raghava: A hybrid approach for predicting promiscuous MHC class I restricted T cell epitopes. *J Biosci* 32, 31-42 (2007)
71. H Singh, GP Raghava: ProPred1: prediction of promiscuous MHC Class-I binding sites. *Bioinformatics* 19, 1009-1014 (2003)
72. H Singh, GP Raghava: ProPred: prediction of HLA-DR binding sites. *Bioinformatics* 17, 1236-1237 (2001)
73. PA Reche, JP Glutting, EL Reinherz: Prediction of MHC class I binding peptides using profile motifs. *Hum Immunol* 63, 701-709 (2002)
74. M Nielsen, O Lund: NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* 10, 296 (2009)



75. J Chen, H Liu, J Yang, KC Chou: Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. *Amino Acids* 33, 423-428 (2007)

76. ZR Yang: Prediction of caspase cleavage sites using Bayesian bio-basis function neural networks. *Bioinformatics* 21, 1831-1837 (2005)

**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>