[Frontiers in Bioscience 4882-4888, May 1, 2008]

Methods and protocols for the assessment of protein allergenicity and cross-reactivity

Joo Chuan Tong[1], Martti Tapani Tammi[2,3,4]

[1]Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613, [2]Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, [3]Department of Biochemistry, National University of Singapore, 8 Medical Drive, Singapore 117597, [4]Karolinska Institutet, Department of Microbiology, Tumor and Cell Biology, Stockholm, Sweden

TABLE OF CONTENTS

1. ABSTRACT

Allergy is a prevalent health problem in developed countries. With advances in genomic and proteomic technologies, there is a rapid increase in allergy-related data, including allergen sequences, allergic cross-reactivity, molecular structures, clinical measurements, and atmospheric concentrations. The more and more complex allergy data is fueling the need for advanced ways in information management and analysis. Computational methods and resources are increasingly the driving force in allergy research. For example, allergen-specific databases are important data sources for allergen characterization. T-cell and B-cell epitope prediction tools focus on identifying immunogenic regions on allergenic proteins. Allergenicity and cross-reactivity prediction tools are increasingly being applied to assess the potential allergenicity of proteins. This review provides an introduction to the growing literature in this area, with particular emphasis on recent developments in bioinformatics relevant to the study of allergens.

2. INTRODUCTION

Allergies are constantly increasing in industrialized countries, affecting more than a quarter of the general population (1, 2). Food-related allergies were reported to affect 4% of the US population (3) and 2.4% within the EU (3-5). Among the children, the incidence of asthma and eczema (atopic dermatitis) were 10% and 15% respectively (6). Allergic diseases are caused by adverse immunological reactions to otherwise innocuous proteins known as allergens. Allergens may be "major" or "minor" depending on whether greater or less than 50% of patients tested react with the corresponding allergen-specific immunoglobulin E (IgE) antibodies in the given test-system (7). Type I hypersensitive reaction is induced when an allergen cross-links IgE antibodies on the mast cells or basophils, resulting in the release of inflammatory mediators (8). This may be followed by a late-phase reaction characterized by the influx of T-cells, eosinophils and monocytes (9). Atopic individuals may develop one or

more manifestations of the disease including asthma, conjunctivitis, dermatitis, rhinitis, and the severe anaphylaxis (3, 8).

The problems associated with allergy and other hypersensitivity reactions catalyzed significant investments and research into new technologies for high-throughput screening of allergy related genes and their products (10-12). As a consequent, there has been an explosion in allergy-related information, taking the form of allergen sequences, allergic cross-reactivity, molecular structures, clinical measurements, and even atmospheric concentrations. Clinical and functional information of common allergens are presently available in various public databases, scientific literature and hospital databases. With such a proliferation of knowledge sources, there is a pressing need for good data management and analysis tools (13).

In time, information technology found its way into making sense of these disparate forms of information. Allergy informatics provides a platform for the integration of basic research and clinical medicine. Major areas of focus include data management and computational analysis. A number of specialized databases were developed for storing allergy-related information with aims of speeding up allergy research (13). T-cell and B-cell epitope prediction tools facilitate the identification of immunogenic regions on allergenic proteins (14-16). Allergenicity and cross-reactivity prediction tools have applications in allergy immunotherapies and vaccine design (17, 18). Each of these topics will be outlined, with an emphasis on the state of the art potentialities of bioinformatics in allergy research as well as on still-open questions.

3. ALLERGEN DATA SOURCES

Data accessibility is critical for detailed characterization and analysis of allergens. Huge amounts of allergen sequences and related information have been accumulated in the literature and case reports. These data are collected and stored in a structured way in a number of databases. Every year, more than a hundred biological databases are described in the Molecular Biology Database Collection (19). The most important allergen resources are reviewed in this section.

3.1. General-purpose databases

General-purpose databases assign biological entities with unique names and characterize their primary sequences or structural features. The international collaborative GenBank (20), DNA Data Bank of Japan (DDBJ) (21) and EMBL (22) serve as worldwide repositories for nucleotide sequences of different origins. The three databases synchronize their records on a daily basis. Swiss-Prot (23) and Protein Information Resource (PIR) (24) provide comprehensive and expertly annotated protein sequence and functional information. A total of 338 protein allergen sequences are currently (June 2007) indexed by Swiss-Prot. TrEMBL (Translated EMBL) has been established as a computer-annotated protein sequence database complementing Swiss-Prot (23). This supplement consists of all translation of EMBL nucleotide sequences that are not available in Swiss-Prot. Protein Data Bank (PDB) (25) is the single worldwide archive of structural data of biological macromolecules. As of June 2007, more than 380 structures related to known allergens have been deposited in PDB.

3.2. Specialized databases

The issue of quality and accessibility of data derived from general-purpose databases is not a recent concern in allergy informatics. Brusic and coworkers (26) raised eight concerns on the consistency and accuracy of data derived from public databases. 1) Allergen sequences in scientific literature may not be submitted to sequence databases. 2) Sequences submitted to public databases may not be released. 3) Variant sequences may be listed only in comments but not as separate entries. 4) Cross-referencing issues with obsolete accession numbers. 5) Duplicate data due to obsolete accessions. 6) Sequence corrections may not be reported to public databases. 7) Lack of synchronization between databases. 8) Incorrect or ambiguous annotations and keywords.

To address these issues, a variety of allergen-specific knowledge sources have been compiled to ensure the relevance and quality of data (26, 27). Historically, the Biotechnology Information for Food Safety (BIFS) database is the first data collection in the field (28). The core of BIFS contains three files: food allergens, non-food allergens, and wheat gluten proteins. The current (June 2007) update contains information on 453 food allergens (64 animals, 389 plants), 645 non-food allergens, and 75 wheat gluten proteins. The Allergen Nomenclature database of the International Union of Immunological Societies (IUIS) (http://www.allergen.org) serves as a central resource for ensuring uniformity and consistency of allergen designations (29). In order to maintain data integrity, the database is curated by committee members and only allergens that can induce IgE-mediated allergy (reactivity > 5%) in humans are included. This property makes IUIS one of the most widely used and authoritative allergen data source. As of June 2007, IUIS contains more than 779 allergens and isoallergens from over 150 species. The Structural Database of Allergen Proteins (SDAP) (30) (http://fermi.utmb.edu/SDAP/) contains information of 887 allergenic proteins. Where available, each SDAP entry is annotated with information such as allergen name, source, sequence, structure, IgE epitopes, literature references, and links to the major public databases. The Food Allergy Research and Resource Program (FARRP) Protein AllergenOnline Database (31) (http://allergenonline.com) contains 1251 sequences of known and putative allergens derived from scientific literature and public databases. The InformAll database (http://www.informall.eu.com/) stores general, biochemical and clinical information on 248 allergenic food materials of both plant and animal origin. The AllAllergy database (http://allallergy.net/) details information of more than 4500 allergenic chemicals and proteins. A comprehensive description on the background information, synonyms, functions and adverse reactions of

each allergen is provided. Allergome (32) (http://www.allergome.org/) emphasizes the annotation of allergens that causing an IgE-mediated disease. The database currently (last update in November 2004) contains information derived from 5800 selected scientific literature. Other data sources exist and have been reviewed elsewhere (26).

4. DATA ANALYSIS TOOLS

4.1. B-cell epitope prediction

The interactions of allergens with IgE antibodies involve binding to antigenic determinants or epitopes on the surface of allergen molecules (33). The antibody binding site is predominantly hydrophobic (34), formed by three hypervariable loops of diverse length and amino acid composition (35). In general, approximately 10% of B-cell epitopes are linear, consisting of a single continuous stretch of amino acids along the polypeptide chain (36). Most epitopes, though, are thought to be conformational, where distantly separated residues of the polypeptide chain are brought into spatial proximity by protein folding (37). Mutational analysis of IgE epitopes have shown that IgE binding could be reduced or eliminated by single-site amino acid substitution (38). Solution structures of antibody-antigen complexes revealed that antibodies with dissimilar binding site structures may exhibit similar specificities for common epitopes (39) and not all residues within an epitope are functionally important for binding (40). As such, the amino acid sequences of B-cell epitopes appear to be weakly conserved and difficult to characterize.

Despite these difficulties, a number of B-cell epitope prediction algorithms have been reported. Most efforts have been focused on screening linear epitopes, partly, due to lower complexity in software development, and also because the experimental design of conformational epitopes is difficult (41). Early approaches relied on the use of propensity scales to create sequence profiles (42-44). It is arguable that such techniques are useful in practice. Blythe and Flower (45) have recently published a careful validation of 484 amino acid propensity scales, and this work gives an astonishing revelation that even the best set of scales and parameters performed only slightly better than random. To the present time, strategies drawn from artificial intelligence have also produced limited success when applied to linear B-cell epitopes (36, 46). However, Kanduc (47) recently introduced the principle of proteomic similarity in defining allergen epitopes, so indicating new avenues for studying allergy phenomena at molecular level. An alternative approach is the use of structural knowledge to guide the design of computational algorithms. The use of structural information parallels the noticeable shift in B-cell epitope prediction methodologies over the past few years, away from sequence-derived properties to much more structure-guided techniques (48-51). This is, in part, the consequence of disappointing predictive performance using sequence data, and also because of the rapidly increasing number of three-dimensional structures of antibody-antigen complexes available in the PDB. The move towards structure-based approaches is also supported by studies demonstrating that both linear and conformational epitopes can be predicted (49-51). Detailed understanding of the structural determinants for antibody-antigen interactions is also useful for the design of hypoallergens for immunotherapeutic vaccines (52).

4.2. T-cell epitope prediction

CD4+ T helper (Th)-cells recognize endogenously processed allergens as short peptide fragments in association with major histocompatibility complex (MHC) class II molecules on antigen-presenting cells (53). Th1 cells produce interferon γ (IFN-γ) and tumor necrosis factor β (TNF-β) and are involved in delayed-type hypersensitivity (DTH) reactions (54, 55). By contrast, Th2 cells produces interleukin 4 (IL-4), IL-5, IL-10 and IL-13, which are responsible for the activation and recruitment of IgE antibody-producing B-cells, mast cells and eosinophils (54, 55). T-cell responses have been studied in patients sensitized to food (56), and allergen-specific T-cell epitopes have been reported (57). The MHC haplotype influences individual immune responses against specific allergens (58). As of June 2007, 548 protein-coding HLA class II alleles had been identified (http://www.anthonynolan.org.uk/HIG/). Binding studies show that each HLA allele recognizes a limited set of peptides. Truncational analysis revealed that class II binding peptides are highly variable in length and the core recognition regions or binding registers are predominantly 9 amino acids long (59). Consequently, the number of potential peptide candidates easily exceeds 10^{11} (60). It is not feasible to experimentally determine the HLA specificities for each and every candidate peptide sequences.

Many methods and programs have been developed and tested in an array of MHC class I and class II alleles. A detailed description of computational strategies for the study of MHC-peptide interactions was given in a recent review (61). These include procedures based on binding motifs (62, 63), binding matrices (64-67), decision trees (68, 69), hidden Markov models (HMMs) (70), support vector machines (SVMs) (71-73), artificial neural networks (ANNs) (74, 75), quantitative structure-activity relationship (QSAR) analysis (76, 77), homology modeling (78, 79), docking (80, 81) and protein threading techniques (82, 83). *Dermatophagoides pteronyssinus* (Der p) 2 is a major source of perennial indoor aeroallergens (84). Analysis of antigenic determinants expressed on the p2 allergen of Der p has revealed that T-cell determinants collectively span the entire length of the molecule (59). Successful application of HLA class II predictive model for identifying the binding registers of Der p2 derived T-cell epitopes has been described (85). There has been increasing focus on the design of computational technologies that allow the prediction of promiscuous peptides that are capable of binding to a wide array of MHC molecules (86, 87). The method enables the design of peptide vaccines with improved global coverage by ensuring that HLA alleles that are present in most individuals from all major ethnic groups bind to at least one of the peptides in the vaccine. Dynamic activities over the past year have also seen at least six reports of algorithms that integrated the different sub-components of the antigen processing

pathway such as the transporter associated with antigen processing (TAP) and proteasomal cleavage specificities (88-91).

4.3. Allergenicity and cross-reactivity prediction

The use of computational models to accelerate assessment of protein allergenicity is a very promising and attractive field. Early ideas in circulation relied on the use of sequence similarity with known allergens to guide the identification of potentially allergenic sequences (92). In 1996, the International Food Biotechnology Council (IFBC) and the Allergy and Immunology Institute of the International Life Sciences Institute (ILIS) developed a decision-tree approach for assessing the allergenic potential of genetically engineered crop plants (92). This approach for assessing protein allergenicity was widely adopted by the agricultural biotechnology industry. Eventually, this was modified by the World Health Organization (WHO) and Food and Agriculture Organization (FAO) in a joint expert consultation on foods derived from biotechnology (92). In the consultation report, guidelines were established for the evaluation of allergenicity of genetically modified foods. In addition to biological tests on the protein of interest, two decision criteria have been proposed for the assessment of allergenic potential. Briefly, a protein is considered allergenic if it possesses an identity of six or more contiguous amino acids or a minimum of 35% sequence similarity over a window of 80 amino acids with a known allergen. Although these approaches have led to the discovery of many new allergens, their inherent limitations are starting to become apparent. Such approaches are reported to be neither specific nor sensitive (93-95), and clinical scientists did not embrace these tools with any real enthusiasm.

Eventually, more sophisticated bioinformatic tools for assessing the allergenic potential of protein sequences have emerged. Computational models based on machine-learning algorithms began to be exploited in allergy research in the early 2000s. Zorzet and coworkers (96) described the use of the FASTA3 algorithm with *k*-Nearest-Neighbour (kNN) classifier for assessing potential food allergenicity of newly introduced proteins. Soeria-Atmadja *et al.* (97) extended the analysis on a larger set of allergens using a combination of kNN classifier, Bayesian linear Gaussian classifier and Bayesian quadratic Gaussian classifier. Cui *et al.* (98) as well as Saha and Raghava (99) reported the use of support vector machine (SVM) for the prediction of novel allergen proteins. An example of the use of allergen-representative peptides to detect potentially allergenic proteins has been reported by Björklund *et al.* (100). The use of Fourier transform to detect compact patterns in allergens was also reported (93). Given the rising concerns of allergy-related problems, it should also be expected that many more advanced methods will appear in the literature.

5. SUMMARY AND OUTLOOK

With the first worked examples appearing more than a decade ago, and a significant number of databases and programs having now reported, allergy informatics is now at a transitional stage of its development. Major concepts including quality data management and analysis appear to have gained wide acceptance. Attention is now focused on deciding how best to use the rich publicly available data sources so that it adds significant value to allergy research. Already, more and more integrated bioinformatic analysis tools are now available in increasing numbers of allergen-specific databases (30, 31). This, together with advances with computational infrastructures, will bring increased focus on the development of computational techniques for large-scale analysis of allergen and allergy-related data, thus facilitating the discovery of new knowledge not possible by traditional experimental techniques alone. The recent finding that to date known allergens have no or few bacterial homologues in contrast to randomly selected control protein sequences is an example of how bioinformatics has led to new ideas in allergy research (101). Together with experimental and clinical research, this approach can significantly accelerate our understanding of the complex allergy pathway.

6. REFERENCES

1. Mekori, Y. A.: Introduction to allergic diseases. *Crit Rev Food Sci Nutr* 36 Suppl, S1-18 (1996)
2. Nieuwenhuizen, N. E. & A. L. Lopata: Fighting food allergy: current approaches. *Ann N Y Acad Sci* 1056, 30-45 (2005)
3. Sampson, H.: Update on food allergy. *J Allergy Clin Immunol* 113, 805-819 (2004)
4. Young, E., M. D. Stoneham, A. Petruckevitch, J. Barton & R. Rona: A population study of food intolerance. *Lancet* 343, 1127-1130 (1994)
5. Cookson, W.: The immunogenetics of asthma and eczema: a new focus on the epithelium. *Nat Rev Immunol* 4, 978-988 (2004).
6. Scientific Committee on Plants: The occurrence of severe food allergies in the EU: Report of experts participating in task 7.2. Scientific Cooperation Program in the EU (1998)
7. Larsen, J. N. & H. Løwenstein: Allergen nomenclature. *J Allergy Clin Immunol* 97, 577-578 (1996)
8. Sutton, B. J. & H. J. Gould: The human IgE network. *Nature* 366, 421-428 (1993)
9. Gould, H., B. Sutton, A. Beavil, R. Beavil, N. McCloskey, H. Coker, D. Fear & L. Smurthwaite: The biology of IGE and the basis of allergic disease. *Ann Rev Immunol* 21, 579-628 (2003)
10. Cookson, W.: Genetics and genomics of asthma and allergic diseases. *Immunol Rev* 190, 195-206 (2002)
11. Deinhofer, K., H. Sevcik, N. Balic, C. Harwanegg, C., R. Hiller, H. Rumpold, M. W. Mueller & S. Spitzauer: Microarrayed allergens for IgE profiling. *Methods* 32, 249-254 (2004).
12. Pineiro, C., J. Barros-Velazquez, J. Vazquez, A. Figueras & J. M. Gallardo: Proteomics as a tool for the investigation of seafood and other marine products. *J Proteome Res* 2, 127-135 (2003)
13. Brusic, V.: Information management for the study of allergies. *Inflamm Allergy Drug Targets* 5, 35-42 (2006)
14. Friedl-Hajek, R., M. D. Spangfort, C. Schou, H. Breiteneder, H. Yssel & R. J. Joost van Neerven:

Identification of a highly promiscuous and an HLA allele-specific T-cell epitope in the birch major allergen Bet v 1: HLA restriction, epitope mapping and TCR sequence comparisons. *Clin Exp Allergy* 29, 478-487 (1999)
15. De Lalla, C., T. Sturniolo, L. Abbruzzese, J. Hammer, A. Sidoli, F. Sinigaglia & P. Panina-Bordignon: Cutting edge: identification of novel T-cell epitopes in Lol p5a by computational prediction. *J Immunol* 163, 1725-1729 (1999).
16. Mueller, G. A., A. M. Smith, M. D. Chapman, G. S. Rule & D. C. Benjamin: Hydrogen exchange nuclear magnetic resonance spectrometry mapping of antibody epitopes on the house dust mite allergen Der p 2. *J Biol Chem* 276, 9359-9356 (2001)
17. Mirza, O., A. Henriksen, H. Ipsen, J. N. Larsen, M. Wissenbach, M. D. Spangfort & M. Gajhede: Dominant epitopes and allergic cross-reactivity: complex formation between a Fab fragment of a monoclonal murine IgG antibody and the major allergen from birch pollen Bet v 1. *J Immunol* 165, 331-338 (2000)
18. Santiago, M. L., J. C. Hafalla, J. D. Kurtis, G. L. Aligui, P. M. Wiest, R. M. Olveda, G. R. Olds, D. W. Dunne & B. L. Ramirez: Identification of the *Schistosoma japonicum* 22.6-kDa antigen as a major target of the human IgE response: similarity of IgE-binding epitopes to allergen peptides. *Int Arch Allergy Immunol* 117, 94-104 (1998)
19. Galperin, M. Y.: The molecular biology database collection: 2007 update. *Nucleic Acids Res* 35, D3-4 (2007)
20. Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, B. A. Rapp & D. L. Wheeler: GenBank. *Nucleic Acids Res* 30, 17-20 (2002)
21. Tateno, Y., T. Imanishi, S. Miyazaki, K. Fukami-Kobayashi, N. Saitou, H. Sugawara & T. Gojobori: DNA Data Bank of Japan (DDBJ) for genome scale research in life science. *Nucleic Acids Res* 30, 27-30 (2002)
22. Kanz, C., P. Aldebert, N. Althorpe, W. Baker, A. Baldwin, K. Bates, *et al.*: The EMBL nucleotide sequence database. Nucleic Acids Res 33, D29-33 (2005)
23. O'Donovan, C., M. J. Martin, A. Gattiker, E. Gasteiger, A. Bairoch & R. Apweiler: High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Brief Bioinform* 3, 275-284 (2002)
24. Wu, C. H., H. Huang, L. Arminski, J. Castro–Alvear, Y. Chen, Z. Z. Hu, *et al.*: The protein information resource: an integrated public resource of functional annotation of proteins. *Nucleic Acids Res* 30, 35-37 (2002)
25. Westbrook, J., Z. Feng, S. Jain, T. N. Bhat, N. Thanki, V. Ravichandran, *et al.*: The protein data bank: unifying the archive. *Nucleic Acids Res* 30, 245-248 (2002)
26. Brusic, V., M. Millor, N. Petrovsky, S. Gendel, O. Gigonzac & S. Stelman: Allergen Databases. *Allergy* 58, 1093-1100 (2003)
27. Gendel, S. M.: Bioinformatics and food allergens. *J AOAC Int* 87, 1417-22 (2004)
28. Gendel, S. M.: Sequence databases for assessing the potential allergenicity of proteins used in transgenic foods. *Adv Food Nutr Res* 42, 63-92 (1998)
29. Hoffman, D., H. Lowenstein, D. Marsh, T. Platts-Mills & W. Thomas: Allergen nomenclature. *Bull World Health Organ* 72, 796-806 (1994)
30. Ivanciuc, O., C. H. Schein & W. Braun: SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res* 31, 359-362 (2003)
31. Hileman, R. E., A. Silvanovich, R. E. Goodman, E. A. Rice, G. Holleschak, J. D. Astwood & S. L. Hefle: Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. *Int Arch Allergy Immunol* 128, 280-291 (2002)
32. Mari, A., V. Mari & A. Ronconi: Allergome - a database of Allergenic molecules: structure and data implementations of a web-based resource *J Allergy Clin Immunol* 115, S87 (2005)
33. Rouvinen, J., J. Rautiainen, T. Virtanen, T. Zeiler, J. Kauppinen, A. Taivainen & R. Mäntyjärvi: Probing the molecular basis of allergy. Three-dimensional structure of the bovine lipocalin allergen Bos d 2. J Biol Chem 274, 2337-2343 (1999)
34. Nair, D. T., K. Singh, N. Sahu, K. V. Rao & D. M. Salunke: Crystal structure of an antibody bound to an immunodominant peptide epitope: novel features in peptide-antibody recognition. *J Immunol* 165, 6949-6955 (2000)
35. Poljak, R. J.: Structure of antibodies and their complexes with antigens *Mol Immunol* 28, 1341-1345 (1991)
36. Larsen, J. E. P., O. Lund & M. Nielsen: Improved method for predicting linear B-cell epitopes. *Immunome Res* 2, 2 (2006)
37. Barlow, D. J., M. S. Edwards & J. M. Thornton: Continuous and discontinuous protein antigenic determinants. *Nature* 322, 747-748 (1986)
38. Helm, R. M., G. Cockrell, C. Connaughton, C. M. West, E. Herman, H. A. Sampson, G. A. Bannon & A. W. Burks: Mutational analysis of the IgE-binding epitopes of P34/Gly m Bd30 K. *J Allergy Clin Immunol* 105, 378-384 (2000)
39. Nair, D. T., K. Singh, Z. Siddigui, B. P. Nayak, K. V. Rao & D. M. Salunke : Epitope recognition by diverse antibodies suggests conformational convergence in an antibody response. J Immunol 168, 2371-2382 (2002)
40. Selo, I., G. Clement, H. Bernard, J. Chatel, C. Creminon, G. Peltre & J. Wal: Allergy to bovine beta-lactoglobulin: specificity of human IgE to tryptic peptides. *Clin Exp Allergy* 29, 1055-1063 (1999)
41. Olive, C., I. Toth & D. Jackson: Technological advances in antigen delivery and synthetic peptide vaccine developmental strategies. *Mini Rev Med Chem* 1: 429-438 (2001)
42. Hopp, T. & K. Woods: Prediction of protein antigenic determinants from amino acid sequence. *Proc Natl Acad Sci U S A* 78, 3824-3828 (1981)
43. Parker, J. M., D. Guo & R. S. Hodges: New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray derived accessible sites. *Biochemistry* 25, 5425-5432 (1986)
44. Van Regenmortel, M. H. & G. Daney de Marcillac: An assessment of prediction methods for locating continuous epitopes in proteins. *Immunol Lett* 17, 95-107 (1988)

45. Blythe, M. J. & D. R. Flower: Benchmarking B cell epitope prediction: underperformance of existing methods. *Protein Sci* 14, 246-248 (2005)
46. Saha, S. & G. P. Raghava: Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 65, 40-48 (2006)
47. Kanduc D: Correlating low-similarity peptide sequences and allergenic epitopes. *Curr Pharm Des*., in press (2007)
48. Haste Andersen, P., M. Nielsen & O. Lund: Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 15, 2558-2567 (2006)
49. Schreiber, A., M. Humbert, A. Benz & U. Dietrich: 3D-Epitope-Explorer (3DEX): localization of conformational epitopes within three-dimensional structures of proteins, *J Comput Chem* 26, 879–887 (2005)
50. Kulkarni-Kale, U., S. Bhosle & A. S. Kolaskar: CEP: a conformational epitope prediction server. *Nucleic Acids Res* 33, W168–W171 (2005)
51. Kolaskar, A. S. & U. Kulkarni-Kale: Prediction of three-dimensional structure and mapping of conformational epitopes of envelope glycoprotein of Japanese encephalitis virus. *Virology* 261, 31-42 (1999)
52. Valenta, R. & D. Kraft: Recombinant allergen molecules: tools to study effector cell activation. *Immunol Rev* 179, 119-127 (2001)
53. Schwartz, R. H.: T-lymphocyte recognition of antigen in association with gene products of the major histocompatibility complex. *Annu Rev Immunol* 3, 237-261 (1985)
54. Romagnani, S.: The Th1/Th2 paradigm. *Immunol Today* 18, 263-266 (1997)
55. Maggi, E.: The Th1/Th2 paradigm in allergy. *Immunoltechnology* 4, 233-244 (1998)
56. Woodfolk, J. A.: T-cell responses to allergens. *J Allergy Clin Immunol* 119, 280-294 (2007)
57. Bohle, B.: T-cell epitopes of food allergens. *Clin Rev Allergy Immunol* 30, 97-108 (2006)
58. Ceppellini, M.: Biological and evolutionary significance of the major histocompatibility complex. *G Batteriol Virol Immunol* 69, 39-48 (1976)
59. Krco, C. J., J. Harders, S. Chapoval & C. S. David: Immune response of HLA-DQ transgenic mice to house dust mite allergen p2: identification of HLA-DQ restricted minimal epitopes and critical residues. *Clin Immunol* 97, 154-161 (2000)
60. Brusic, V. & N. Petrovsky: Immunoinformatics and its relevance to understanding human immune disease. *Expert Rev Clin Immunol* 1, 145-157 (2005)
61. Tong, J. C., T. W. Tan & S. Ranganathan: Methods and protocols for prediction of immunogenic epitopes. *Brief Bioinform* 8, 96-108 (2006)
62. Falk, K., O. Rötzschke, S. Stevanovic, G. Jung & H. G. Rammensee: Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 351, 290-296 (1991)
63. Parker, K. C., M. A. Bednarek & J. E. Coligan: Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J Immunol* 152, 163-175 (1994)
64. Davenport, M. P., I. A. P. Ho Shon & A. V. S. Hill: An empirical method for the prediction of T-cell epitopes. *Immunogenetics* 42, 392-397 (1995)
65. Gulukota, K., J. Sidney, A. Sette & C. DeLisi: Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *J Mol Biol* 267, 1258-1267 (1997)
66. Reche, P. A., J. P. Glutting & E. L. Reinherz: Prediction of MHC class I binding peptides using profile motifs. *Hum Immunol* 63, 701-709 (2002)
67. Guan, P., I. A. Doytchinova, C. Zygouri & D. R. Flower: MHCPred: bringing a quantitative dimension to the online prediction of MHC binding. *Appl Bioinformatics* 2, 63-66 (2003)
68. Savoie, C. J., N. Kamikawaji, T. Sasazuki & S. Kuhara: Use of BONSAI decision trees for the identification of potential MHC class I peptide epitope motifs. *Pac Symp Biocomput.*, 182-189 (1999)
69. Segal, M. R., M. P. Cummings & A. E. Hubbard: Relating amino acid sequence to phenotype: analysis of peptide-binding data. *Biometrics* 57, 632-642 (2001)
70. Mamitsuka, H: Predicting peptides that bind to MHC molecules using supervised learning of hidden Markov models. *Proteins* 33, 460-474 (1998)
71. Zhao, Y., C. Pinilla, D. Valmori, R. Martin & R. Simon: Application of support vector machines for T-cell epitopes prediction. *Bioinformatics* 19, 1978-1984 (2003)
72. Dönnes, P. & A. Elofsson: Prediction of MHC class I binding peptides, using SVMHC. *BMC Bioinformatics* 3, 25 (2002)
73. Bhasin, M. & G. P. S. Raghava: SVM based method for predicting HLA-DRB1*0401 binding peptides in an antigen sequence', *Bioinformatics* 20, 421-423 (2004)
74. Brusic, V., G. Rudy, M. Honeyman, J. Hammer & L. Harrison: Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics* 14, 121-130 (1998)
75. Nielsen, M., C. Lundegaard, P. Worning, S. L. Laumøller, K. Lamberth, S. Buus, S. Brunak & O. Lund: Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci* 12, 1007-1117 (2003)
76. Doytchinova, I. A., V. Walshe, P. Borrow & D. R. Flower: Towards the chemometric dissection of peptide—HLA-A*0201 binding affinity: comparison of local and global QSAR models. *J Comput Aided Mol Des* 19, 203-212 (2005)
77. Doytchinova, I. A., P. Guan & D. R. Flower: Quantitative structure-activity relationships and the prediction of MHC supermotifs. *Methods* 34, 444-453 (2004)
78. Michielin, O., I. Luescher & M. Karplus: Modeling of the TCR-MHC-peptide complex. *J Mol Biol* 300, 1205-1235 (2000)
79. Logean, A., A. Sette & D. Rognan: Customized versus universal scoring functions: application to class I MHC-peptide binding free energy predictions. *Bioorg Med Chem Lett* 11, 675-679 (2001)
80. Bordner, A. J. & R. Abagyan: Ab initio prediction of peptide-MHC binding geometry for diverse class I MHC allotypes. *Proteins* 63, 512-526 (2006)

81. Tong, J. C., T. W. Tan & S. Ranganathan: Modeling the structure of bound peptide ligands to major histocompatibility complex. *Protein Sci* 13, 2523-2532 (2004)
82. Altuvia, Y., A. Sette, J. Sidney, S. Southwood & H. Margalit: A structure-based algorithm to predict potential binding peptides to MHC molecules with hydrophobic binding pockets. *Hum Immunol* 58, 1-11 (1997)
83. Schueler-Furman, O., R. Elber & H. Margalit: Knowledge-based structure prediction of MHC class I bound peptides: a study of 23 complexes. *Fold Des* 3, 549-564 (1998)
84. Steward, G. A.: Dust mite allergens. *Clin Rev Allergy Immunol* 13, 135-150 (1995)
85. Tong, J. C., G. L. Zhang, T. W. Tan, J. T. August, V. Brusic & S. Ranganathan: Prediction of HLA-DQ3.2β ligands: Evidence of multiple registers in class II binding peptides. *Bioinformatics* 22, 1232-1238 (2006)
86. Brusic, V., N. Petrovsky, G. Zhang & V. B. Bajic: Prediction of promiscuous peptides that bind HLA class I molecules. *Immunol Cell Biol* 80, 280-285 (2002)
87. Lund, O., M. Nielsen, C. Kesmir, A. G. Petersen, C. Lundegaard, P. Worning, C. Sylvester-Hvid, K. Lamberth, G. Røder, S. Justesen, S. Buus & S. Brunak: Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics* 12, 797-810 (2004)
88. Doytchinova, I. A. & D. R. Flower: Class I T-cell epitope prediction: improvements using a combination of proteasome cleavage, TAP affinity, and MHC binding. *Mol Immunol* 43, 2037-2044 (2006)
89. Doytchinova, I. A., P. Guan & D. R. Flower: EpiJen: a server for multistep T cell epitope prediction. *BMC Bioinformatics* 7, 131 (2006)
90. Tenzer, S., B. Peters, S. Bulik, O. Schoor, C. Lemmel, M. M. Schatz, P. M. Kloetzel, H. G. Rammensee, H. Schild & H. G. Holzhutter: Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci* 62, 1025-1037 (2005)
91. Donnes, P. & O. Kohlbacher: Integrated modeling of the major events in the MHC class I antigen processing pathway. *Protein Sci* 14, 2132-2140 (2005)
92. FAO/WHO: Codex principles and guidelines on foods derived from biotechnology. Joint FAO/WHO food standards programme, Rome, Italy (2003)
93. Li, G. B., P. Issac & A. Krishnan: Predicting allergenic proteins using wavelet transform. *Bioinformatics* 20, 2572-2578 (2004)
94. Hileman, R. E., A. Silvanovich, R. E. Goodman, E. A. Rice, G. Holleschak, J. D. Astwood & S. L. Hefle: Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. *Int Arch Allergy Immunol* 128, 280-291 (2002)
95. Stadler, M. B. & B. M. Stadler: Allergenicity prediction by protein sequence. *FASEB J* 17, 1141-1143 (2003)
96. Zorzet, A., M. Gustafsson & U. Hammerling: Prediction of food protein allergenicity: a bioinformatic learning systems approach. In Silico Biol 2, 525-534 (2002)
97. Soeria-Atmadja, D., A. Zorzet, M. G. Gustafsson & U. Hammerling: Statistical evaluation of local alignment features predicting allergenicity using supervised classification algorithms. *Int Arch Allergy Immunol*, 133, 101-112 (2004)
98. Cui, J., L. Y. Han, H. Li, C. Y. Ung, Z. Q. Tang, C. J., Zheng, Z. W. Cao & Y. Z. Chen: Computer prediction of allergen proteins from sequence-derived protein structural and physicochemical properties. *Mol Immunol* 44, 514-520 (2007)
99. Saha, S. & G. P. S. Raghava: AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res* 34, W202-209 (2006)
100. Björklund, A. K., D. Soeria-Atmadja, A. Zorzet, U. Hammerling & M. G. Gustafsson: Supervised identification of allergen-representative peptides for in silico detection of potentially allergenic proteins. *Bioinformatics*, 21, 39-50 (2005)
101. Emanuelsson, C. & M. D. Spangfort: Allergens as eukaryotic proteins lacking bacterial homologues. *Mol. Immunol.*, 44, 3256-3260 (2007)

Key Words: allergen data, bioinformatics, immunologic databases

Send correspondence to: Martti T. Tammi, Department of Biological Sciences, National Umiversity of Singapore, 14 Science Drive 4, 117543 Singapore, Tel: 65-6516-4255, Fax: 65-6779-5671, E-mail: martti@nus.edu.sg

http://www.bioscience.org/current/vol13.htm