

BIOINFORMATICS: MANAGING THE KNOWLEDGE ON ALLERGENS

A. Mari, MD (1)

Molecular biology and all the improved biochemical and immunochemical techniques have lead to the identification of a number of allergens, increasing our knowledge on them (1-4). The time trend of these historical changes can be easily summarized by considering the number of publication dedicated to allergens in the last 40 years. They are now around 6,000, half of which have been published in the last 5 years. At the same time we have seen the development of new tools for diagnosis linked to the nanotechnology. Nanotechnology-based systems applied to IgE sensitizing/triggering compounds is a novel application of proteomic microarrays. They can be based on allergenic extracts (5-10), but more recently native and recombinant purified allergens have been used (11-14). The use of microarrays in allergy is expected to change the approach to epidemiology and diagnosis of IgE-mediated diseases. Combining the use of allergenic molecules and microarrays will lead to a wealth of data on allergic diseases, that will require managing and interpreting.

Internet adoption, and in a broader sense information technology, has allowed us to create online databases and tools that can be easily reached by worldwide users.

In the last two decades information and data have been collected from available sources in order to create database of scientific knowledge. For instance, a growing list of molecular biology databases is reported every year (15). A database is defined as capable to acquire, store, visualize, analyze and interpret data. Allergy databases available in Internet are reported in table 1. The main features of each database are given in a comparative way as they are on the web or described in scientific journal (16-20). Web sites considered deal with protein antigens recognized by human IgE. Only free access non-commercial

resources are considered. Although each database follows a specific philosophy determined by the task, most of them are dedicated to collect allergen molecular structures. These allergen lists are used to feed algorithms employed for defining allergenicity of new compounds. Anyway allergen sequence databases are the more numerous. They have been mainly set up for computational purposes rather than to create general resources for allergy knowledge.

The potential application of information technology to the allergy field should also consider two new fields: literature mining and data mining. There is a critical need for extraction, integration and management from sometimes highly heterogeneous data (21). There is now a specific research field dedicated to develop informatics tools for literature mining (22). These tools should allow us to summarize major themes and topics, and reveal trends in references. The increasing trend of biomedical publications recorded in the allergy field will ask to develop specific tools, either literature databases or literature miners that are specific for the allergy field. The analysis of current web-based resources shows us that only two of the web sites report citations for each allergen. This way to present citations does not offer a cross sectional view of literature data. The rapidly expanding fields of allergen knowledge is prompting us to develop new computational methods for managing, analyzing and visualizing literature data (22;23).

This approach should contribute to optimizing the exploitation of scientific productions by providing investigators and clinicians with an easy interface with complex and large literature resources. There is a trend in using Internet for managing general medical information. This information technology approach to medical data managing results in a simple and quick way for patient/physician interaction, in optimization of health care costs for patients and healthcare providers, and in a reduction of errors (24). It is now evident that storing of this mass of information is creating a mine of real time clinical

(1) Allergy Data Laboratories s.c. - Via Malipiero 28 I - 04100 Latina (Italy) - adriano.mari@allergome.org

Web Site	URL	Data Source	Allergen Sources	Allergen Molecules	Allergen Sequence Database	Computational Tools	Biochemistry Molecular Biology	Clinical / Epidemiological Data	Internal Search Engine	Internal Data Cross-linking	User Alert	References	Last Update
Official Allergen Nomenclature of WHO-IUIS (*)	www.allergen.org	Submission	No	All	No	No	+	No	No	No	No	+	Feb 02, 2005
Allergen Database (ADB)	www.csl.gov.uk/allergen	Sequence Databases	No	All	External Links	No	+	No	Yes	No	No	+	Given for each allergen
Allergen Online (FARRP) *	www.allergenonline.com	Sequence Databases	No	All	Yes & Ext. Links	Yes & External	++	No	Yes	No	Yes	No	Oct 1, 2004
Allergome	www.allergome.org/	Literature & Sequence Databases	All	All	External Links	No	+++	Yes	Yes	Yes	Yes	+++	Mar 12, 2005
AllerMatch *	www.allermatch.org/	Sequence Databases	No	All	Yes &	Yes	++	No	No	No	No	No	Dec 10, 2004
AllerPredict	http://sdmc.i2r.astar.edu.sg/Templar/DB/Allergen/	Sequence Databases	No	All	Yes & Ext. Links	Yes & External	+	No	Yes	Yes	No	+	Given for each allergen
Bioinformatics for Food Safety (BIFS)	http://www.iit.edu/~sgendel/fa.htm	Sequence Databases	No	All	External Links	No	No	No	No	No	No	+	Not given
InformAll (Protall)	http://foodallergens.itf.ac.uk/	Literature & Sequence Databases	Food	Food	External Links	No	+++	Yes	Yes	No	No	+++	Jan 18, 2005
Structural Database of Allergen Proteins (SDAP) *	http://fermi.utmb.edu/SDAP/sdap_src.html	Sequence Databases	No	All	Yes & Ext. Links	Yes & External	+++	No	Yes	No	Yes	No	Given for each allergen
WebAllergen *	http://weballergen.bii.a-star.edu.sg/	Sequence Databases	No	All	Yes	Yes	No	No	No	No	No	No	Not given

Table 1: Main features of Internet-available resources on allergens.

(+) An arbitrary grading of the supported computational tools, and other supplied information is indicated.

(*) Official web site of the WHO/IUIS allergen nomenclature sub-committee.

data. Allergies are the most common diseases worldwide and allergen-based diagnostic and epidemiological data are available from published studies. The use of IT in allergy data mining should be evaluated in the near future. Considering the source of data (skin tests, IgE testing, provocation tests) a great effort should be put in standardizing data source for data entry. Highly purified allergenic molecules coupled on microarrays along with fully automated laboratory systems should lead us to the real time managing of allergy data. Molecular biology, nanotechnology and IT are offering us a great opportunity to adapt and transform the use of currently available tools including pioneering web-based databases. Computational tools have to be applied to this dynamic mass of information to extract, interpret and visualize relevant new findings. Curing of data by human experts to define decision points will anyway be unavoidable.

Key-words: Allergy, allergen, database, Internet, sequence, computational, Information technology.

Notes and Acknowledgments: Data from Internet databases have been retrieved on May 2005. The Allergome platform is an independent project managed and financially supported by Allergy Data Laboratories s.c., a non-for-profit company. The project is also supported by unrestricted grants from Pharmacia Diagnostics (Uppsala, Sweden), Indoor Biotechnologies (Cardiff, United Kingdom), Laboratorios Leti (Barcelona, Spain), Diagnostic Product Corporation (Los Angeles, CA, USA) Stallergenes (Antony, France), Allergopharma (Reinbek, Germany), VBC-Genomics (Vienna, Austria).

References

1. Kazemi-Shirazi L., Niederberger V., Linhart B., Lidholm J., Kraft D., Valenta R. - Recombinant marker allergens: diagnostic gatekeepers for the treatment of allergy. *Int Arch Allergy Immunol* 2002;127(4):259-68.
2. Chapman M.-D. - Allergen nomenclature. *Clin Allergy Immunol* 2004;18:51-64.
3. Breiteneder H., Radauer C. - A classification of plant food allergens. *J Allergy Clin Immunol* 2004;113(5):821-30.
4. Ferreira F., Hawranek T., Gruber P., Wopfner N., Mari A. - Allergic cross-reactivity: from gene to the clinic. *Allergy* 2004;59(3):243-67.
5. Su X., Chew F.-T., Li S.-F. - Piezoelectric quartz crystal based label-free analysis for allergy disease. *Biosens Bioelectron* 2000;15(11-12):629-39.
6. Wiltshire S., O'Malley S., Lambert J., Kukanskis K., Edgar D., Kingsmore S.-F. et al. - Detection of multiple allergen-specific IgEs on microarrays by immunoassay with rolling circle amplification. *Clin Chem* 2000;46(12):1990-3.
7. Mullenix M.-C., Wiltshire S., Shao W., Kitos G., Schweitzer B. - Allergen-specific IgE detection on microarrays using rolling circle amplification: correlation with in vitro assays for serum IgE. *Clin Chem* 2001;47(10):1926-9.
8. Kim T.-E., Park S.-W., Cho N.-Y., Choi S.-Y., Yong T.-S., Nahm B.-H. et al. - Quantitative measurement of serum allergen-specific IgE on protein chip. *Exp Mol Med* 2002;34(2):152-8.
9. Bacarese-Hamilton T., Mezzasoma L., Ingham C., Ardizzone A., Rossi R., Bistoni F. et al. - Detection of allergen-specific IgE on microarrays by use of signal amplification techniques. *Clin Chem* 2002;48(8):1367-70.
10. Lebrun S.-J., Petchpud W.-N., Hui A., McLaughlin C.-S. - Development of a sensitive, colorimetric microarray assay for allergen-responsive human IgE. *J Immunol Methods* 2005;300(1-2):24-31.
11. Hiller R., Laffer S., Harwanegg C., Huber M., Schmidt W.-M., Twardosz A. et al. - Microarrayed allergen molecules: diagnostic gatekeepers for allergy treatment. *FASEB J* 2002;16(3):414-6.
12. Suck R., Nandy A., Weber B., Stock M., Fiebig H., Cromwell O. - Rapid method for arrayed investigation of IgE-reactivity profiles using natural and recombinant allergens. *Allergy* 2002;57(9):821-4.
13. Jahn-Schmid B., Harwanegg C., Hiller R., Bohle B., Ebner C., Scheiner O. et al. - Allergen microarray: comparison of microarray using recombinant allergens with conventional diagnostic methods to detect allergen-specific serum immunoglobulin E. *Clin Exp Allergy* 2003;33(10):1443-9.
14. Shreffler W.-G., Beyer K., Chu T.-H., Burks A.-W., Sampson H.-A. - Microarray immunoassay: Association of clinical history, in vitro IgE function, and heterogeneity of allergenic peanut epitopes. *J Allergy Clin Immunol* 2004;113(4):776-82.
15. Galperin M.-Y. - The Molecular Biology Database Collection: 2005 update. *Nucleic Acids Res* 2005;33(Database issue):D5-D24.
16. Gendel S.-M. - Sequence databases for assessing the potential allergenicity of proteins used in transgenic foods. *Adv Food Nutr Res* 1998;42:63-92.
17. Ivanciuc O., Schein C.-H., Braun W. - SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res* 2003;31(1):359-62.
18. Mari A., Riccioli D. - The Allergome Web Site - A Database of Allergenic Molecules. Aim, Structure, and Data of a web-based resource. *J Allergy Clin Immunol* 2004;113(2):S301.
19. Fiers M.-W., Kleter G.-A., Nijland H., Peijnenburg A.-A., Nap J.-P., Van Ham R.-C. - Allermatch TM, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC Bioinformatics* 2004;5(1):133.
20. Riaz T., Hor H.-L., Krishnan A., Tang F., Li K.-B. - WebAllergen: a web server for predicting allergenic proteins. *Bioinformatics* 2005;21(10):2570-1.
21. Haynes R.-B., McKibbin K.-A., Wilczynski N.-L., Walter S.-D., Werre S.-R. - Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. *BMJ* 2005;330(7501):1179.
22. Chaussabel D. - Biomedical literature mining: challenges and solutions in the 'omics' era. *Am J Pharmacogenomics* 2004;4(6):383-93.
23. Chaussabel D., Sher A. - Mining microarray expression data by literature profiling. *Genome Biol* 2002;3(10):RESEARCH0055.
24. Oliven A., Michalake I., Zalman D., Dorman E., Yeshurun D., Odeh M. - Prevention of prescription errors by computerized, on-line surveillance of drug order entry. *Int J Med Inform* 2005;74(5):377-86.