

Allergen databases—A critical evaluation

Since the 1980s, a large number of allergenic proteins have been identified and characterized. Consequently, several institutions and research groups established databases to collect the available but scattered information. These databases contain overlapping data, but address different user groups, list allergens based on varying criteria, and may provide additional tools such as sequence comparisons. This editorial provides an overview of the strengths and weaknesses of the currently most widely used, freely accessible databases of IgE-binding allergens. A summary of the basic features and a critical evaluation of the databases are shown in Tables 1 and 2.

The World Health Organization/International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature database,¹ established in 2000, is maintained by the WHO/IUIS Allergen Nomenclature Sub-Committee, an international body of currently 22 leading experts in molecular allergology. This database provides a systematic and unambiguous nomenclature for proteins that induce IgE-mediated allergies in humans. Allergens are included after a detailed review by the Allergen Nomenclature Sub-Committee.² Each record includes basic data on biochemical properties, sequences, and allergenicity. The database serves as a reference for researchers, clinicians, regulatory authorities, and the industry. Most other

TABLE 1 Allergen databases discussed in this article

Name	Maintained by	URL	Update frequency
Actively updated allergen sequence databases			
WHO/IUIS Allergen Nomenclature Database	WHO/IUIS Allergen Nomenclature Sub-Committee	www.allergen.org	Continuous
AllergenOnline (FARRP Allergen Database)	Food Allergy Research and Resource Program, Department of Food Science and Technology, University of Nebraska-Lincoln, Lincoln, NE, USA	www.allergenonline.org	Annual
Comprehensive Protein Allergen Resource (COMPARE)	Protein Allergens, Toxins and Bioinformatics Committee, Health and Environmental Sciences Institute	comparedatabase.org	Annual
Allergome	Allergy Data Laboratories, Latina, Italy	www.allergome.org	Continuous
AllerBase	Bioinformatics Centre, Savitribai Phule Pune University, India	bioinfo.net.in/AllerBase/Home.html	Weekly
Inactive but still accessible allergen sequence databases			
Structural Database of Allergenic Proteins (SDAP)	Sealy Center for Structural Biology, Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX, USA	fermi.utmb.edu/SDAP/	Latest update: Feb. 25, 2013
InformAll Allergenic Food Database	Manchester Institute of Biotechnology, Manchester, UK	research.bmh.manchester.ac.uk/informall/allergenic-foods/	Latest update: Oct. 18, 2006
Allergen-related databases			
Immune Epitope Database (IEDB)	National Institute for Allergy and Infectious Diseases, Bethesda, MD, USA	iedb.org	Quarterly
AllFam	Division of Medical Biotechnology, Institute of Pathophysiology and Allergy Research, Medical University of Vienna, Austria	www.meduniwien.ac.at/allfam/	Annual

TABLE 2 Evaluation of allergen databases

Database	Strengths	Weaknesses
WHO/IUIS Allergen Nomenclature Database	- Official reference for allergen names and corresponding sequences - New allergens are included only after detailed review by the Allergen Nomenclature Sub-Committee	- Slow updating of existing entries - Limited number of literature references
AllergenOnline (FARRP Allergen Database)	- List of allergens reviewed by a panel of experts - Each record contains literature references that justify its inclusion in the database - Database available for download	- Minimal data provided apart from sequence and route of exposure - Caution: the database also contains IgE-binding antigens from parasites not involved in allergic reactions
Comprehensive Protein Allergen Resource (COMPARE)	- Peer-reviewed database of allergen sequences - Allergen identification process fully disclosed on the website - Database available for download	- No reference to official WHO/IUIS allergen names - No allergy-related data included - Sequence accession numbers not linked to a database - Caution: the database also contains IgE-binding antigens from parasites not involved in allergic reactions
Allergome	- Comprehensive collection of data on allergen sources and allergens, obtained from other databases and the literature - Extensive list of publications included in each record	- All potential allergens with published data are included without filtering - Caution: not all used IUIS-like allergen names are officially recognized
AllerBase	- Data on allergens, IgE epitopes, and allergen-specific antibodies - Compilation of allergen data from various databases and the literature - Allergen records include experimental data with associated publications, grouped by method	Inclusion criteria of non-IUIS allergens not specified
Structural Database of Allergenic Proteins (SDAP)	- Peer-reviewed allergen sequence database - Collection of bioinformatics tools - Extended data on allergen structures and epitopes	- Inclusion criteria for allergens not specified - No updates since 2013
InformAll Allergenic Food Database	Data on allergenic foods and food allergens	No updates since 2006
Immune Epitope Database (IEDB)	Comprehensive collection of data on B- and T-cell epitopes	Complex user interface requires some effort to retrieve specific information
AllFam	Classification of allergens into protein families	Some AllFam families are broadly defined and do not reflect potential cross-reactivity

allergen databases use allergen data recorded in the WHO/IUIS Allergen Nomenclature database as their main source.

AllergenOnline was established in 2005 by the Food Allergy Research and Resource Program in the Department of Food Science and Technology at the University of Nebraska in Lincoln, Nebraska, USA, to provide a peer-reviewed database of allergen sequences.³ It is used by the agricultural industry for the allergenicity assessment of proteins planned to be introduced into genetically modified crops. For this purpose, sequence similarity search tools are provided. Putative allergen sequences are collected from the National Center for Biotechnology Information (NCBI) database by searching for "allerg*," supplemented by data from the WHO/IUIS Allergen Nomenclature and Allergome⁴ databases and filtered, based on peer-reviewed publications, by a panel of expert reviewers.

The Comprehensive Protein Allergen Resource (COMPARE) is a database of allergen sequences created as a tool for food safety assessment similar to AllergenOnline. It was first published in 2017

by the Health and Environmental Sciences Institute, an international consortium comprising academia, government, industry, and non-government organizations. It provides an annually updated freely downloadable list of allergen sequences that is created by first identifying putative allergen sequences in the NCBI protein database by an automated search using keyword-based filters, as detailed on the COMPARE website. Genuine allergens are then identified by a panel of peer reviewers. Database records contain sequences, accession numbers, and key publications, but no allergy-related data.

The Allergome database,⁴ released in 2003, is maintained by Allergen Data Laboratories, a company located in Italy. Allergome houses the most comprehensive collection of data on allergen sources and allergens, and is useful for scientists, clinicians, and the industry. Nevertheless, users are advised to use Allergome with caution, as virtually all IgE-binding proteins are included without filtering for clinical relevance. Data are compiled from the literature and from other databases. Each allergen record comprises various

biochemical and clinical data, links to other databases and an extensive list of literature references grouped by topic. The website also contains a collection of tools for data analysis and visualization.

AllerBase⁵ was created to integrate data from allergen, sequence, epitope, antibody, and literature databases into a single platform. Allergen records contain links to many other databases. Experimental data which the inclusion of a specific allergen was based on are grouped by method and supplemented with associated literature references. This recently created database provides a useful resource for researchers, clinicians, and the industry.

The Structural Database of Allergenic Proteins (SDAP)⁶ was created in 2002 by the Sealy Center for Structural Biology, Department of Biochemistry and Molecular Biology at the University of Texas Medical Branch in Galveston, Texas. It hosts data on allergen sequences, epitopes, and experimentally determined structures as well as a large number of homology models. Database records are linked to a collection of bioinformatics tools for sequence analysis and comparison. Allergen data were compiled from the WHO/IUIS Allergen Nomenclature Database and supplemented by allergens retrieved from sequence, structure, and literature databases after being reviewed by an external scientific advisory board. SDAP was regularly updated until 2013.

The InformAll food allergen database⁷ was set up within the framework of an EU-funded project. It provided data on allergenic foods and food allergens intended for the general public, food allergic consumers, agro-food industry, health professionals, and regulators. The database has not been updated since 2006, and hence, many of the provided links are dysfunctional.

The Immune Epitope Database (IEDB)⁸ was released in 2006 by the National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA. It hosts data on experimentally determined B-cell and T-cell epitopes in the context of infectious diseases, allergy, autoimmunity, and transplantation. Data are submitted by researchers or extracted from the literature using a combination of automated searches in PubMed. Retrieved publications are manually curated by a board of expert reviewers following detailed published criteria. IEDB's sophisticated user interface allows for targeted searches for specific information and provides a useful tool for researchers interested in epitopes of allergens, including carbohydrate epitopes such as α -Gal.

AllFam⁹ was established in 2007 by the Department of Pathophysiology and Allergy Research of the Medical University of Vienna, Austria. AllFam provides an easy-to-use interface for the classification of allergens into protein families in order to meet the demands of many researchers and clinicians. Such an evolutionary classification aids in the prediction of cross-reactivity and provides insights into factors that make proteins allergenic. AllFam is based on data from the WHO/IUIS Allergen Nomenclature database and AllergenOnline and the protein family definitions of the Pfam database (pfam.xfam.org).

Different databases cater for the needs of diverse user groups by either focusing on specific types of data (eg, molecular or clinical) or aiming to provide comprehensive information. A metadatabase

that encompasses all available allergen-related information does not exist yet, although Allergome and AllerBase were developed with that goal in mind. In any case, all databases face increasing challenges of data curation and updating, website maintenance, and financing.

ACKNOWLEDGMENTS

Author HB acknowledges the support of the Austrian Science Fund (FWF) Doctoral Program MCCA W1248-B30.

CONFLICT OF INTEREST

The authors are members of the WHO/IUIS Allergen Nomenclature Sub-Committee, which maintains the IUIS allergen nomenclature database, and members of the AllFam team.

KEYWORDS

allergen databases, allergen informatics, allergen nomenclature, food safety, IgE-mediated allergies

Christian Radauer 
 Heimo Breiteneder 

Institute of Pathophysiology and Allergy Research, Medical University of Vienna, Vienna, Austria

Correspondence

Christian Radauer and Heimo Breiteneder, Institute of Pathophysiology and Allergy Research, Medical University of Vienna, Vienna, Austria.
 Emails: christian.radauer@mv.ac.at; heimo.breiteneder@mv.ac.at

ORCID

Christian Radauer  <https://orcid.org/0000-0001-9920-4449>
 Heimo Breiteneder  <https://orcid.org/0000-0003-2022-8689>

REFERENCES

1. Pomés A, Davies JM, Gadermaier G, et al. WHO/IUIS Allergen Nomenclature: providing a common language. *Mol Immunol*. 2018;100:3-13.
2. Goodman RE, Breiteneder H. The WHO/IUIS Allergen Nomenclature. *Allergy*. 2019;74(3):429-431.
3. Goodman RE, Ebisawa M, Ferreira F, et al. AllergenOnline: a peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. *Mol Nutr Food Res*. 2016;60(5):1183-1198.
4. Mari A, Rasi C, Palazzo P, Scala E. Allergen databases: current status and perspectives. *Curr Allergy Asthma Rep*. 2009;9(5):376-383.
5. Kadam K, Karbhal R, Jayaraman VK, Sawant S, Kulkarni-Kale U. AllerBase: a comprehensive allergen knowledgebase. *Database (Oxford)*. 2017;2017:bax066.
6. Ivanciuc O, Schein CH, Braun W. SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res*. 2003;31(1):359-362.

7. Mills EN, Jenkins JA, Sancho AI, et al. Food allergy information resources for consumers, industry and regulators. *Arb Paul Ehrlich Inst Bundesamt Sera Impfstoffe Frankf A M*. 2006;95:17-25.
8. Vita R, Mahajan S, Overton JA, et al. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res*. 2019;47(D1):D339-D343.
9. Radauer C, Bublin M, Wagner S, Mari A, Breiteneder H. Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *J Allergy Clin Immunol*. 2008;121(4):847-852.

How to cite this article: Radauer C, Breiteneder H. Allergen databases—A critical evaluation. *Allergy*. 2019;74:2057–2060. <https://doi.org/10.1111/all.13841>