Current perspectives

Nomenclature and structural biology of allergens

Martin D. Chapman, PhD,[a] Anna Pomés, PhD,[a] Heimo Breiteneder, PhD,[b] and Fatima Ferreira, PhD[c] *Charlottesville, Va, and Vienna and Salzburg, Austria*

Purified allergens are named using the systematic nomenclature of the Allergen Nomenclature Sub-Committee of the World Health Organization and International Union of Immunological Societies. The system uses abbreviated Linnean genus and species names and an Arabic number to indicate the chronology of allergen purification. Most major allergens from mites, animal dander, pollens, insects, and foods have been cloned, and more than 40 three-dimensional allergen structures are in the Protein Database. Allergens are derived from proteins with a variety of biologic functions, including proteases, ligand-binding proteins, structural proteins, pathogenesis-related proteins, lipid transfer proteins, profilins, and calcium-binding proteins. Biologic function, such as the proteolytic enzyme allergens of dust mites, might directly influence the development of IgE responses and might initiate inflammatory responses in the lung that are associated with asthma. Intrinsic structural or biologic properties might also influence the extent to which allergens persist in indoor and outdoor environments or retain their allergenicity in the digestive tract. Analyses of the protein family database suggest that the universe of allergens comprises more than 120 distinct protein families. Structural biology and proteomics define recombinant allergen targets for diagnostic and therapeutic purposes and identify motifs, patterns, and structures of immunologic significance. (J Allergy Clin Immunol 2007;119:414-20.)

Key words: *Allergen, nomenclature, IgE, asthma, protein families, allergic disease, protein structure, biologic function*

The biochemistry of allergens is underpinned by a Linnean system of nomenclature that is maintained by the World Health Organization (WHO) and International Union of Immunological Societies (IUIS) Allergen Nomenclature Sub-Committee. The systematic nomenclature was the brainchild of the late Dr David Marsh (Johns Hopkins University), who authored a seminal chapter on ''Allergens and the genetics of allergy'' in the 1970s. This chapter reviewed allergen structure, immune response, and immunogenetics and also provided the first definitions of major and minor allergens.[1] At that time, allergens were described using a variety of generic names, such as Antigen E, Rye 1, and Cat-1, and it was not uncommon for researchers to use different names for the same allergen. In 1980, Marsh, together with Dr Henning Lowenstein and Dr Thomas Platts-Mills, developed the systematic nomenclature during the 13th Symposium of the Collegium Internationale Allergologicum (Lake Bodensee, Germany). A committee, including Drs Te Piao King and Larry Goodfriend, drafted the nomenclature and developed criteria for biochemical properties and allergenic importance that would qualify allergens in the new system. The systematic nomenclature was adopted by the WHO/IUIS and published in the Bulletin of the WHO in 1986 and in revised form in 1994.[2-5] Allergens are named using the first 3 letters of the genus, followed by a single letter for the species and a number indicating the chronologic order of allergen purification. Thus the major cat allergen (formerly Cat-1) became *Felis domesticus* allergen 1 or Fel d 1.

The systematic allergen nomenclature proved to be robust and accommodated the explosion of data on new allergens that occurred in the 1980s and 1990s, when the most important allergens from mites, animal dander, insects, pollens, molds, and foods were cloned. Allergens entered into the nomenclature are being used to develop allergen-specific diagnostics and to formulate recombinant allergen vaccines.[6-8] Allergen biochemistry is now entering a new era of structural biology and proteomics that will require sophisticated tools for data processing and bioinformatics and might require further

From [a]Indoor Biotechnologies, Inc, Charlottesville; [b]the Medical University of Vienna; and [c]Christian Doppler Laboratory for Allergy Diagnosis and Therapy, University of Salzburg.

H. B. is supported by Austrian Science Fund Grant SFB F01802. F. F. is supported by the Christian Doppler Research Association, the Austrian Science Fund (Grants S8802 and P16456), and the Oesterreichische Nationalbank (Grant no. 10150).

Disclosure of potential conflict of interest: M. D. Chapman owns stock in Indoor Biotechologies, Inc; has received grant support from the National Institute of Environmental Health Sciences Small Business Innovation Research award program; and is employed by Indoor Biotechnologies, Inc. A. Pomés is employed by Indoor Biotechnologies, Inc. F. Ferreira has consulting arrangements with Indoor Biotechnologies, Inc, and has received grant support from Biomay AG. H. Breiteneder has declared that he has no conflict of interest.

Received for publication October 17, 2006; revised October 31, 2006; accepted for publication November 2, 2006.

Available online December 14, 2006.

Reprint request: Martin D. Chapman, PhD, INDOOR Biotechnologies, Inc, 1216 Harris St, Charlottesville, VA 22903. E-mail: mdc@inbio.com.

0091-6749/$32.00

doi:10.1016/j.jaci.2006.11.001

Abbreviations used
EF hand: Calcium-binding motif (E-helix-loop-F-helix) in a "hand" configuration
IUIS: International Union of Immunological Societies
PR-10: Pathogenesis-related group 10
WHO: World Health Organization

delineation of the nomenclature. Increasingly, the wealth of structural information is enabling the biologic function of allergens to be established and the assignment of allergen function to diverse protein families. In this article we review the allergen nomenclature system and recent advances in structural biology that have established the form and function of many important allergenic proteins.

CURRENT ALLERGEN NOMENCLATURE

The current allergen nomenclature was developed through 2 iterations in 1986 and 1994, since which it has been unchanged.[2-5] The nomenclature is not italicized, has a space after each of the first two elements, and uses Arabic numerals: hence Der p 1, Bet v 1, Fel d 1, and Amb a 1, for example. The nomenclature covers different molecular forms of the same allergen: isoallergens and isoforms (or variants). Isoallergens are multiple molecular forms of the same allergen that share extensive IgE cross-reactivity. They are defined in the nomenclature as allergens from a single species with 67% or greater amino acid sequence identity. The most prolific example is birch pollen allergen, Bet v 1, which has more than 40 sequences representing 31 isoallergens, showing 73% to 98% sequence identity.[9] The Bet v 1 isoallergens are distinguished by additional numbers: Bet v 1.01 through Bet v 1.31. Similarly, 4 isoallergens of ragweed allergen, Amb a 1, are listed as Amb a 1.01, Amb a 1.02, Amb a 1.03, and Amb a 1.04. The terms *isoform* or *variant* refer to polymorphic variants of the same allergen, which typically show greater than 90% sequence identity. Isoforms are distinguished in the nomenclature by 2 additional numbers. The 42 isoforms of Bet v 1 are listed as Bet v 1.0101, Bet v 1.0102, Bet v 1.0103, and so on. Recent studies have shown that mite allergen sequences derived from environmental isolates by means of high-fidelity PCRs show extensive numbers of isoforms: 23 for Der p 1 (Der p 1.0101 to Der p 1.0123) and 13 for Der p 2.[10-12] These polymorphisms might affect T-cell responses or alter antibody-binding sites and should be taken into account in designing allergen formulations for immunotherapy.[12]

The reader is referred to a recent review for finer points of the current nomenclature.[9] To submit a newly defined allergen, investigators should download the "New Allergen Name" form from the official website of the WHO/IUIS Sub-Committee on Allergen Nomenclature at www.allergen.org. The application is reviewed by the Allergen Nomenclature Sub-Committee, which is chaired by Dr Heimo Breiteneder (Medical University of Vienna, Vienna, Austria) and comprises 19 experts in the field (see Table E1 in the Online Repository at www.jacionline.org). The molecular properties of allergens to be included in the nomenclature must be unambiguously defined by submitting nucleotide and amino acid sequence data, by intrinsic molecular properties (molecular weight, isoelectric point, and secondary structure), by purification of the allergen to homogeneity, and by monospecific antibodies. The importance of the allergen in causing IgE responses should be demonstrated by *in vitro* testing, by biologic testing (histamine release or skin testing), and by comparing the prevalence of IgE antibody binding in a large group of allergic patients.[9] The goal of the Allergen Nomenclature Subcommittee is simply to provide systematic nomenclature and clear identification of allergens and not to grade allergens on their importance or assign any ownership rights. Allergens must be shown to cause IgE antibody production in at least 5 individuals to be included, but otherwise, researchers must demonstrate the merits and significance of their particular protein.

ADVANCES IN STRUCTURAL BIOLOGY AND PROTEOMICS

Molecular cloning and searches of GENBANK, EMBL, and other protein databases have allowed the biologic function of many allergens to be assigned based on their amino acid sequence homology to proteins of known function. Assignments based on sequence homology do not prove that an allergen has a given function but do provide evidence that can be used to investigate whether a particular allergen has the putative biologic activity. For example, the homology of Der p 1 to papain and actinidin strongly suggested that Der p 1 was a cysteine protease. This was later confirmed by using functional assays and by x-ray crystallography, which determined the structures of the proenzyme and mature forms of the allergen.[13,14] The Protein Database contains more than 40 three-dimensional structures of allergens. Structural studies often reveal features of biologic importance that might not be apparent from biologic assays.

Allergens belong to protein families with diverse biologic functions that can be summarized as follows:

(1) indoor allergens: enzymes (especially proteases), ligand-binding proteins or lipocalins, albumins, tropomyosins, and calcium-binding proteins;
(2) pollen allergens: pathogenesis-related proteins, calcium-binding proteins, pectate lyases, β-expansins, and trypsin inhibitors; and
(3) plant and animal food allergens: lipid transfer proteins, profilins, seed storage proteins, and tropomyosins.

Indoor allergens

In some cases the biologic function of an allergen might have direct effects on IgE responses and have