# 3

# Allergen Nomenclature

**MARTIN D. CHAPMAN**

*INDOOR Biotechnologies, Inc., and University of Virginia, Charlottesville, Virginia, U.S.A.*



## I. HISTORICAL INTRODUCTION

As with most biochemical disciplines, the history of allergen nomenclature dates back to the time when allergens were fractionated using a variety of "classical" biochemical separation techniques and the active (most allergenic) fraction was usually named according to the whim of the investigator. For allergens, this dates to the 1940s through the late 1950s, when early attempts were made to purify pollen and house dust allergens using phenol extraction, salt precipitation, and electrophoretic techniques. In the early 1960s, ion exchange and gel filtration media were introduced and ragweed "antigen E" was the first allergen to be purified. This allergen, named by King and Norman, was one of five precipitin lines (labeled A–E) that reacted with rabbit polyclonal antibodies to ragweed in Ouchterlony immunodiffusion tests. Following purification, precipitin line E, or "antigen E" was shown to be a potent allergen (1). Later, Marsh, working in Cambridge, England, isolated an important allergen from rye grass pollen (*Lolium perenne*) and used the name "Rye 1" to indicate that this was the first allergen purified from this species (2). In the 1970s, the field advanced

apace and many allergens were purified from ragweed, rye grass, insect venoms, and other sources. The field was led by the laboratory of the late Dr. David Marsh, who had moved to Johns Hopkins University in Baltimore, Maryland. There ragweed allergens Ra3, Ra4, Ra5, and Ra6 and rye grass allergens Rye 2 and Rye 3 were isolated and used for immunological and genetic studies of hay fever. At the same time, Ohman identified a major cat allergen (Cat-1) (3) and Elsayed purified allergen M from codfish (4).

The state of the art in the early 1970s was reviewed in a seminal chapter by Marsh in *The Antigens* (ed. Michael Sela), which described the molecular properties of allergens, the factors that influenced allergenicity, the immune response to allergens, and immunogenetic studies of IgE responses to purified pollen allergens (5). This chapter provided the first clear definition of a "major" allergen, which Marsh defined as a highly purified allergen that induced immediate skin test responses in >90% of allergic individuals—this in contrast to a "minor" allergen, to which <20% of patients had skin test responses. A less stringent standard was subsequently adopted, and today a major allergen is defined as one to which >50% of allergic patients react.

With the introduction of crossed immunoelectrophoresis (CIE) and crossed radioimmunoelectrophoresis (CRIE) for allergen identification by Lowenstein and colleagues in Scandinavia, there was a tremendous proliferation of the number of antigenic proteins and CIE/CRIE peaks identified as allergens (6). Typically, 10 to 50 peaks could be detected in a given allergen based on reactivity with rabbit polyclonal antibodies or IgE antibodies. These peaks were given a plethora of names such as Dp5, Dp42, Ag12, etc. Inevitably, this led to the same allergens being referred to by different names in different laboratories. Thus, mite antigen $P_1$ was also known as Dp42 or Ag12. It was clear that a unified nomenclature was urgently needed.

## A. Three Men in a Boat

The origins of the systematic allergen nomenclature can be traced to a meeting among Drs. David Marsh (at that time at Johns Hopkins University, Baltimore), Henning Lowenstein (at that time at the University of Copenhagen, Denmark) and Thomas Platts-Mills (at that time at Clinical Research Centre, Harrow, UK) on a boat ride on Lake Boedensee, Konstanz, Germany, during the 13th Symposium of the Collegicum Internationale Allergologicum in July 1980 (7). The idea was simply to develop a systematic nomenclature based on the Linnean system, with numerals used to indicate different allergens. It was decided to adopt a system whereby the allergen was described based on the first three letters of the genus and the first letter of the species (in italics) and then by a Roman numeral to indicate the allergen in the chronological order of purification. Thus, ragweed antigen E became *Ambrosia artemisifolia* allergen I or *Amb a* I, and Rye 1 became *Lolium perenne* allergen I or *Lol p* I.

An allergen nomenclature subcommittee was formed under the auspices of the World Health Organization (WHO) and the International Union of Immunological Societies (IUIS), and criteria for including allergens in the systematic nomenclature were established. These included strict criteria for biochemical purity, as well as criteria for determining the allergenic activity of the purified protein. A committee chaired by Marsh and including Lowenstein, Platts-Mills, Dr. Te Piao King (Rockefeller University, New York), and Dr. Larry Goodfriend (McGill University, Canada) prepared a list of allergens that fulfilled the inclusion criteria and established a process for investigators to submit names of newly identified allergens. The original list, published in the *Bulletin of the*

*World Health Organization* in 1986, included 27 highly purified allergens from grass, weed and tree pollens, and house dust mites (8).

The systematic allergen nomenclature was quickly adopted by allergy researchers and proved to be a great success. It was logical, easily understood, and readily assimilated by allergists and other clinicians who were not directly involved with the nitty-gritty of allergen immunochemistry. The nomenclature *Der p* I, *Fel d* I, *Lol p* I, *Amb a* I was used at scientific meetings and in the literature, and expanded rapidly to include newly isolated allergens.

## II. THE REVISED ALLERGEN NOMENCLATURE

### A. Allergens

The widespread use of molecular cloning techniques to identify allergens in the late 1980s and 1990s led to an exponential increase in the number of allergens described. A large number of allergen nucleotide sequences were generated from cDNA- or PCR-based sequencing, and it soon became apparent that the use of Roman numerals (e.g., *Lol p* I through *Lol p* XI) was unwieldy (9–11). The use of italics to denote a purified protein was inconsistent with nomenclature used in bacterial genetics and the HLA system, where italicized names denote a gene product and roman typeface indicates an expressed protein. In 1994 the allergen nomenclature was revised so that the allergen phenotype was shown in roman type and arabic numerals were adopted. Thus *Amb a* I, *Lol p* I, and *Der p* I in the original 1986 nomenclature are referred to as Amb a 1, Lol p 1, and Der p 1 in the current nomenclature, which has been published in several scientific journals (12–14).

#### 1. Inclusion Criteria

A key part of the systematic WHO/IUIS nomenclature is that the allergen should satisfy biochemical criteria, which define the molecular structure of the protein, and immunological criteria, which define its importance as an allergen. Originally, the biochemical criteria were based on establishing protein purity (e.g., by SDS-PAGE, IEF, or HPLC and physicochemical properties including MW, pI, and N-terminal amino acid sequence) (8). Nowadays, the full nucleotide or amino acid sequence is generally required. An outline of the inclusion criteria is shown in Table 1. An important aspect of these criteria is that

**Table 1** Allergens: Criteria for Inclusion in the WHO/IUIS Nomenclature

1. The molecular and structural properties should be clearly and unambiguously defined, including:
   - Purification of the allergen protein to homogeneity.
   - Determination of molecular weight, pI, and carbohydrate composition.
   - Determination of nucleotide and/or amino acid sequence.
   - Production of monospecific or monoclonal antibodies to the allergen.
2. The importance of the allergen in causing IgE responses should be defined by:
   - Comparing the prevalence of serum IgE antibodies in large population(s) of allergic patients. Ideally, at least 50 or more patients should be tested.
   - Demonstrating biological activity, e.g., by skin testing or histamine release assay.
   - Investigating whether depletion of the allergen from an allergic extract (e.g., by immunoabsorption) reduces IgE binding activity.
   - Demonstrating, where possible, that recombinant allergens have comparable IgE antibody binding activity to the natural allergen.

they should provide a "handle" whereby other investigators can identify the same allergen and make comparative studies. Originally, this was achieved by purifying the protein, developing monospecific or monoclonal antibodies to it, and providing either the allergen or antibodies to other researchers for verification. Nucleotide and amino acid sequencing unambiguously identifies the allergen and enables sequence variation between cDNA clones of the same allergen to be defined (15,16). Allergen preparations, sequences, and antibodies submitted for inclusion in the systematic nomenclature are expected to be made available to other investigators for research studies.

A second set of inclusion criteria is based on demonstrating the allergenic activity of the purified allergen, both in vitro and in vivo. Researchers use a variety of techniques for measuring IgE antibodies in vitro, including radioallergosorbent (RAST)-based techniques, CIE/CRIE, radioimmunoassays using labeled allergens, enzyme immunoassay (ELISA), and immunoblotting. These techniques differ in sensitivity, and their efficacy may be affected by a variety of factors. For example, CIE/CRIE is dependent on the quality of polyclonal rabbit antisera. Immunoblotting, which has largely replaced CIE techniques, relies on the allergen being resistant to heating in detergents used for electrophoresis. Whatever technique is used, it is important to screen a large number of sera from an unselected allergic population to establish the prevalence of reactivity. Ideally, 50 or more sera should be screened, although allergens can be included in the nomenclature if the prevalence of IgE reactivity is >5% and they elicit IgE responses in as few as five patients (Table 1,12). "Chimeric" ELISA systems are now available that allow a large number of sera to be screened for IgE antibodies to specific allergens. The assays use a captured monoclonal antibody to bind allergen. Serum IgE antibodies that bind to the allergen complex are detected by biotinylated anti-IgE (Fig. 1). The assay is quantitated using a chimeric mouse anti–Der p 2 and human IgE epsilon antibody and provides results in nanograms per milliliter of allergen-specific IgE. Chimeric ELISA for measuring IgE antibody to Der p 1, Der p 2, and Fel d 1 correlate with Pharmacia CAP measurements and provide useful tools for comparing the prevalence of IgE to specific allergens (17,18).

It is often easier to isolate sequences from cDNA libraries and screen them against panels of sera than it is to work with patients themselves! However, demonstrating that the allergen has biological activity in vivo is critical, especially since many allergens are now produced as recombinant molecules before the natural allergen is purified (if ever). Several mite, cockroach, and fungal allergens (e.g., *Aspergillus, Alternaria, Cladosporium*) have been defined solely using recombinant proteins, and it is unlikely, in most cases, that much effort will be directed toward isolating the natural allergens (9–11,15,16). In these cases, the allergenic activity of the bacterial or yeast expressed recombinant protein should be confirmed in vivo by quantitative skin testing or in vitro by histamine release assays. Skin testing studies have been carried out using a number of recombinant allergens, including Bet v 1, Asp f 1, Bla g 4, Bla g 5, Der p 2, Der p 5, and Blo t 5. These allergens have shown very good biological activity using picogram amounts of proteins.

## 2. Resolving Ambiguities in Nomenclature

Every system has its faults, and allergen nomenclature is no exception. Early on it was recognized that because the system had Linnaean roots, some unrelated allergens would have the same name: *Candida* allergens could be confused with dog allergen (*Canis domesticus*), there are multiple related species of *Vespula* (Vespid) allergens, and *Periplaneta americana* (American cockroach) allergen needs to be distinguished from *Persea americana* (avocado)! These ambiguities have been overcome by adding an additional letter to either the genus or

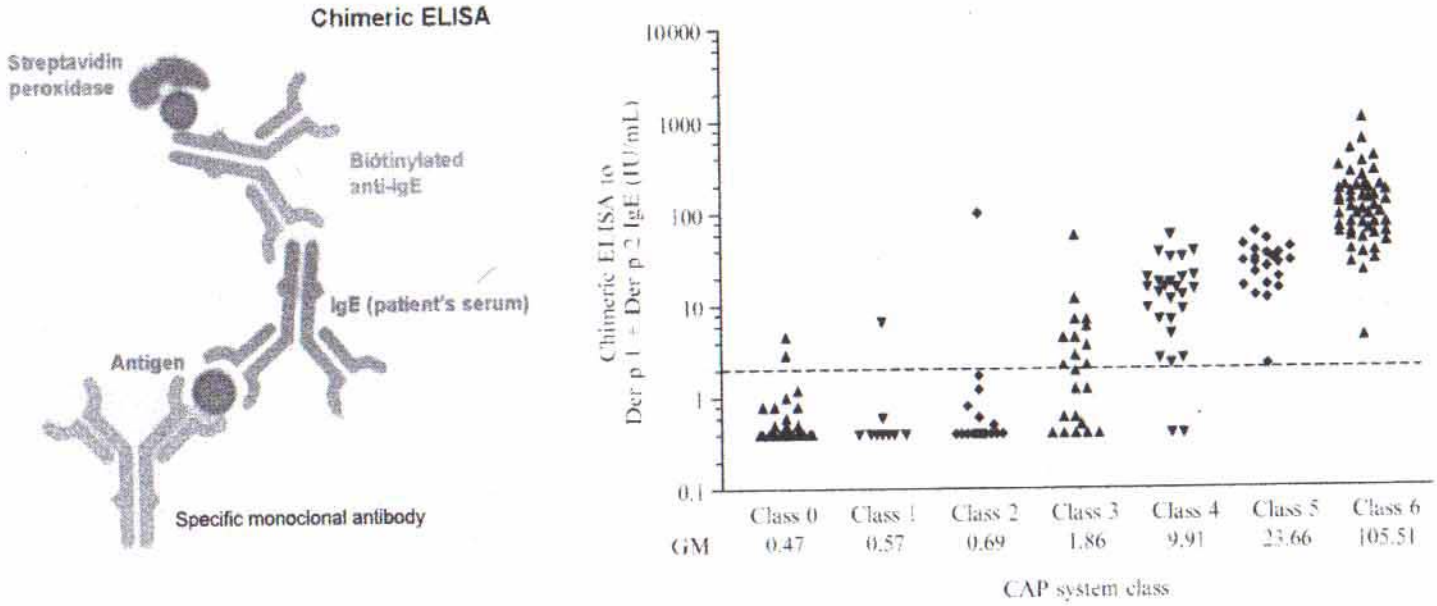


**Figure 1** Chimeric ELISA for measuring allergen-specific IgE. *A*: Schematic graphic of the ELISA. Microtiter plates are coated with monoclonal antibodyfollowed by the relevant allergen and incubated with patient's serum. IgE antibodies that bind to the allergen complex are detected using biotinylated anti-IgE and streptavidin peroxidase. A chimeric IgE anti–Der p 2 is used to generate a control curve, and IgE values for patient's serum are interpolated from this curve. *B:* Correlation between the chimeric ELISA for IgE antibody to Der p 1 and Der p 2 and the Pharmacia CAP system for measuring IgE to house dust mite. Chimeric ELISA values for IgE anti–Der p 1 and IgE anti–Der p 2 were summed and compared with the CAP system. Sera were obtained from 212 patients with asthma, wheezing, and/or rhinitis. There was an excellent quantitative correlation between the chimeric ELISA and CAP ($r = 0.86$, $p < 0.001$). (Reproduced from Trombone et al., Clin Exp Allergy 32:1323–1328, 2002, with permission.)

species name. The preceding examples thus become Cand a 1 (*C. albicans* allergen 1); Ves v 1 or Ves vi 1, to indicate *V. vulgaris* and *V. vidua* allergens, respectively; and Per a 1 and Pers a 1 for the cockroach and avocado allergens. Dog allergen is referred to as Can f 1, from *Canis familiaris*.

Many allergens have biochemical names that describe their biological function and may precede the allergen nomenclature. Examples include egg allergens (ovomucoid and ovalbumin), insect allergens (phospholipase As and hyaluronidases), and tropomyosins from shrimp, mite, and cockroach. In fact, it is common to be able to designate allergens to particular protein families based on sequence homology searches, which have provided important clues to their biological function. Allergens may be enzymes, e.g., proteases (Der p 1, Der p 3, Der p 9) or glutathione transferases (Der p 8, Bla g 5); ligand binding proteins (Bla g 4, Rat n 1, Can f 1, Bos d 2); storage proteins (peanut, Ara h 1); hemoglobins (midge, Chi t 1); plant pathogenesis–related proteins (Bet v 1); or have as yet undetermined functions (mite Group 5 and Group 7 allergens, Group 1 and Group 5 grass pollen allergens). Although several mite and fungal allergens are proteolytic enzymes, the dog allergen Can f 1 has 60% homology to human Van Ebner's gland protein (VEGH), which is a cysteine protease inhibitor. A cystatin allergen (Fel d 3) has also been cloned from a cat skin cDNA library. Fel d 3 has a conserved cysteine protease inhibitor motif that is partially preserved in Can f 1, a lipocalin (Fig. 2) (19). In the allergy literature, it is preferable to use the systematic allergen nomenclature. However, in other contexts, such as comparisons of biochemical activities or protein structure, it may be appropriate or more useful to use the biochemical names. A selected list of the allergen nomenclature and biochemical names of inhalant, food, and venom allergens is shown in Table 2.

The use of molecular cloning has led to the rapid identification of allergen sequences, and multiple allergens have been cloned from several sources. Six or more allergens have been defined from each of the following sources: mite (*Dermatophagoides*), grass and ragweed pollen, cockroach, *Aspergillus, Alternaria*, and latex (Table 2). Homologous allergens have also been cloned from related species, and this can create problems for naming the homologues or unrelated allergens from other species. Mite is a good example. Structural homologues of *Dermatophagoides* allergens have been cloned from *Euroglyphus maynei* (Eur m 1), *Lepidoglyphus destructor* (Lep d 2), and *Blomia tropicalis*

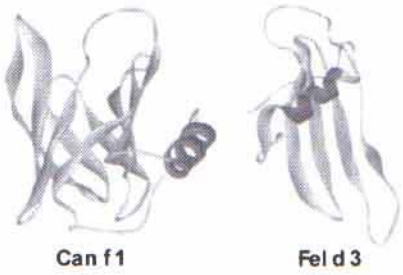


**Figure 2** Molecular modeling of the three-dimensional structures of Can f 1 and Fel d 3, which are thought to function as cysteine protease inhibitors. Fel d 3 has a cysteine protease inhibitor motif (QVVAG) that is located at the tip of the central loop at the bottom of the figure. Similar residues are located in the flattened loop region at the base of the Can f 1 structure. These loop regions are thought to bind to cysteine proteases and inhibit their activity. (Fel d 3 structure reproduced with permission from Clin Exp Allergy 31:1279–1286, 2001.)

**Table 2** Molecular Properties of Common Allergens

| Source | Allergen | MW(kDA) | Homology/function |
|---|---|---|---|
| **Inhalants** | | | |
| **Indoor** | | | |
| House dust mite (*Dermatophagoides pteronyssinus*) | Der p 1 | 25 | Cysteine protease[b] |
| | Der p 2 | 14 | Epididymal protein?[b] |
| | Der p 3 | 30 | Serine protease |
| | Der p 5 | 14 | Unknown |
| Cat (*Felis domesticus*) | Fel d 1 | 36 | (Uteroglobin)[b] |
| Dog (*Canis familiariss*) | Can f 1 | 25 | Cysteine protease inhibitor?[b] |
| Mouse (*Mus musculariss*) | Mus m 1 | | Lipocalin (territory marking protein |
| Rat (*Rattus norvegicus*) | Rat n 1 | 21 | Pheromone-binding lipocalin[b] |
| Cockroach (*Blattella germanica*) | Bla g 2 | 36 | Inactive aspartic protease |
| **Outdoor** | | | |
| Pollen—grasses | | | |
| Rye (*Lolium perenne*) | Lol p 1 | 28 | Unknown |
| Timothy (*Phleum pratense*) | Phl p 5 | 32 | Unknown |
| Bermuda (*Cynodon dactylon*) | Cyn d 1 | 32 | Unknown |
| Weeds | | | |
| Ragweed (*Artemisia artemisifolia*) | Amb a 1 | 38[a] | Pectate lyase[b] |
| | Amb a 5 | 5 | Neurophysins[b] |
| Trees | | | |
| Birch (*Betula verucosa*) | Bet v 1 | 17 | Pathogenesis-related protein[b] |
| **Foods** | | | |
| Milk | β-Lactolobulin | 36 | Retinol-binding[a,b] protein (calycin)[b] |
| Egg | Ovomucoid | 29 | Trypsin inhibitor |
| Codfish (*Gadus callarias*) | Gad c 1 | 12 | Ca-binding protein (muscle parvalbumin) |
| Peanut (*Arachis hypogea*) | Ara h 1 | 63 | Vicilin (seed-storage protein)[b] |
| **Venoms** | | | |
| Bee (*Apis melifera*) | Api m 1 | 19.5 | Phospholipase $A_2$[b] |
| Wasp (*Polestes annularis*) | Pol a 5 | 23 | Mammalian testis proteins |
| Hornet (*Vespa crabro*) | Ves c 5 | 23 | Mammalian testis proteins |
| Fire ant (*Solenopsis invicta*) | Sol i 2 | 13 | Unknown |
| **Fungi** | | | |
| *Aspergillus fumigatus* | Asp f 1 | 18 | Cytotoxin (mitogillin) |
| *Alternaria alternata* | Alt a 1 | 29 | Unknown |
| **Latex** | | | |
| *Hevea brasiliensis* | Hev b 1 | 58 | Elongation factor |
| | Hev b 5 | 16 | Unknown—homologous to kiwi fruit protein of unknown function |

[a] Most allergens have a single polypeptide chain; dimers are indicated.
[b] Allergens of known three-dimensional structure are also indicated.

(Blo t 5), which show >40% homology to the *Dermatophagoides* allergens (11). The problem comes in numbering other allergens cloned from *Lepidoglyphus* or *Blomia* cDNA libraries that may be unrelated to *Dermatophagoides* allergens. Calling the allergen, for example, Blo t 3, in the absence of evidence that *Blomia* produces a homologous allergen to Der p 3, would cause complications if such a homologue were identified at a later date. In these cases, it may be better to use Blo t 11, for example, for the *Blomia* allergen, reserving numbers 1–10 for any allergens related to *Dermatophagoides* that may subsequently be identified.

## B. Isoallergens, Isoforms and Variants

Originally, isoallergens were broadly defined by Marsh and others as multiple molecular forms of the same allergen, sharing extensive antigenic (IgE) cross-reactivity. The revised nomenclature defines isoallergens as allergens from a single species, with similar molecular size, identical biological function, and ≥67% amino acid sequence identity (8). Some allergens that were previously "grandfathered" into the nomenclature as separate entities share extensive sequence homology and some antigenic cross-reactivity, but are named independently and are not considered to be isoallergens. Examples include Lol p 2 and Lol p 3 (65% homology), and Amb a 1 and Amb a 2 (65% homology). The word "group" is now being used more often to describe structurally related allergens from different species within the same genus, or from closely related genera. In these cases, the levels of amino acid sequence identity can range from as little as 40% to ~90%. Similarities in tertiary structure and biological function are also taken into account in describing allergen groups. Examples include the Group 2 mite allergens (Der p 2, Der f 2 and Lep d 2, Gly d 2 and Tyr p 2), showing 40% to 88% homology, and the Group 5 ragweed allergens (Amb a 5, Amb t 5, and Amb p 5), showing ~45% homology. The *Dermatophagoides* Group 2 allergen structures have been determined by X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR). The structures of the Group 2 allergens from other species were modeled on the *Dermatophagoides* structures (Fig. 3). This enabled the structural basis for antigenic relationships between members of the group to be defined (20–22).

The term "variant" or "isoform" is used to indicate allergen sequences that show a limited number of amino acid substitutions (i.e., polymorphic variants of the same allergen). Typically, variants may be identified by sequencing several cDNA clones of a given allergen. Variants have been reported for Der p 1, Der p 2, Amb a 1, Cry j 1, and for the most prolific Bet v 1, for which 42 sequences have been deposited in the GenBank database. Isoallergens and variants are denoted by the addition of four numeral suffixes to the allergen name. The first two numerals distinguish isoallergens and the last two distinguish variants. Thus, for ragweed Amb a 1, which occurs as four isoallergens, showing 12% to 24% difference in amino acid sequence, the nomenclature is as follows:

*Allergen*: Amb a 1
*Isoallergens*: Amb a 1.01, Amb a 1.02, Amb a 1.03, Amb a 1.04

Three variants of each isoallergen occur, showing >97% sequence homology:

*Isoforms*: Amb a 1.0101, Amb a 1.0102, Amb a 1.0103
Amb a 1.0201, Amb a 1.0202, Amb a 1.0203, etc.

Examples showing precisely how the nomenclature for isoforms of mite Group 2 allergens and for the Group 1 allergens of cockroach have been published (20,23). The

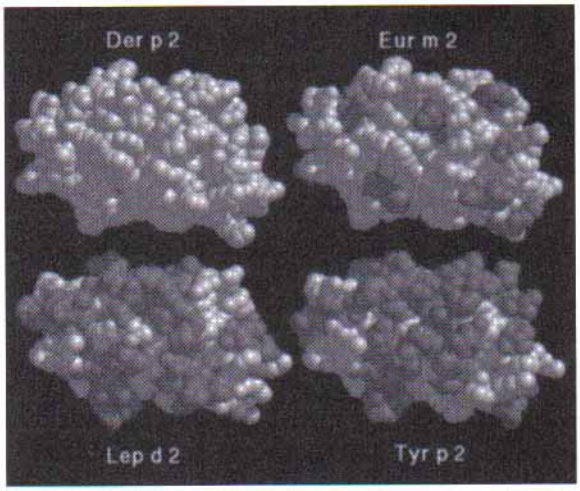


**Figure 3** Space-filling models of Group 2 allergens from house dust mite. Amino acid substitutions are shown in gray scale. The space-filling model of Der p 2 was generated from nuclear magnetic resonance spectroscopy studies and has subsequently been confirmed by X-ray crystallography (22). Eur m 2 shows 85% sequence identity with Der p 2, and seven of the substituted amino acids are shown in gray on the surface structure. There is extensive cross-reactivity between Der p 2 and Eur m 2. In contrast, Lep d 2 and Tyr p 2 show only 40% amino acid identity with the other Group 2 allergens. They show many substitutions on the antigenic surface of the molecules and show limited antigenic cross-reactivity for mAb and human IgE. (Reproduced from Smith et al., J Allergy Clin Immunol 107:977–984, 2001, with permission.)

Group 1 allergens from tree pollen have an unusually high number of isoallergens and variants. The 42 Bet v 1 sequences are derived from 31 isoallergens, which show from 73% to 98% sequence homology and are named Bet v 1.0101 through Bet v 1.3101. The Group 1 allergen from hornbeam (*Carpinus betulus*), Car b 1, has three isoallergens that show 74% to 88% homology (Car b 1.01, 1.02, and 1.03), and the nomenclature committee's most recent records show 15 sequences of Car b 1. Ten variants of hazel pollen allergen, Cor a 1, have also been recorded. The reasons the Group 1 tree pollen allergens have so many variants are unclear. Latex provides another example of distinctions in nomenclature. Hevein is an important latex allergen, designated Hev b 6, which occurs as a 20-kDa precursor with two fragments derived from the same transcript. These moieties are all variants of Hev b 6 and are distinguished as Hev b 6.01 (prohevein, 20-kDa precursor), Hev b 6.02 (5-kDa hevein), and Hev b 6.03 (a 14-kDa C-terminal fragment).

## III. NOMENCLATURE FOR ALLERGEN GENES AND RECOMBINANT OR SYNTHETIC PEPTIDES

In the revised nomenclature, italicized letters are reserved to designate allergen genes. Two genomic allergen sequences have been determined from animal dander allergens: cat allergen, Fel d 1, and mouse urinary allergen, Mus m 1. Fel d 1 has two separate genes encoding chain 1 and chain 2 of the molecule, which are designed Fel d 1A and Fel d 1B, respectively (24). Genomic sequences of Bet v 1, Cor a 1, and apple allergen, Mal d 1, have also been determined.

When recombinant allergens were introduced, researchers often used the term "native allergen" to distinguish the natural protein from the recombinant allergen. However, because "native" has implications for protein structure (i.e., native conformation), it was decided that the term "natural allergen" should be used to indicate any allergen purified from natural source material. Natural allergens may be denoted by the prefix "n" to distinguish them from recombinant allergens, which are identified by the prefix "r" before the allergen name (e.g., nBet v 1 and rBet v 1). There is no distinction between recombinant allergens produced in bacterial, yeast, or mammalian expression systems. Synthetic peptides are identified by the prefix "s", with the particular peptide residues indicated in parentheses after the allergen name. Thus, a synthetic peptide encompassing residues 100–120 of Bet v 1.0101 would be denoted as sBet v 1.0101 (100–120). At this point, the nomenclature, while technically sound, begins to become cumbersome and rather long-winded for most purposes. Additional refinements to the nomenclature cover substitutions of different amino acid residues within synthetic peptides. This aspect of the nomenclature (which is based on that used for synthetic peptides of immunoglobulin sequences) is detailed in the revised nomenclature document, to which aficionados are referred for full details (8).

## IV. THE IUIS SUBCOMMITTEE ON ALLERGEN NOMENCLATURE

Allergens to be considered for inclusion in the nomenclature are reviewed by an IUIS subcommittee, which is currently chaired by Dr. Wayne Thomas, Institute for Child Health, Western Australia, and has eight members (Table 3). The committee meets annually at an international allergy/immunology meeting and discusses new proposals it has received during the year, together with any proposed changes or additions to the nomenclature. There is also a committee-at-large, which is open to any scientist with an interest in allergens, to whom decisions made by the subcommittee are circulated. The procedure for submitting candidate names for allergens to the subcommittee is straightforward. Having purified the allergen and demonstrated its allergenicity, investigators should download the "new allergen name" form from the nomenclature subcommittee Web site (www.allergen.org) and send the completed form to the subcommittee prior to publishing articles describing the allergen. The subcommittee will provisionally accept the author's

**Table 3** The IUIS Subcommittee on Allergen Nomenclature, 2003–2005

| Name | Institution | Country |
|---|---|---|
| Wayne R. Thomas, Ph.D. (chairman) | Western Australia Institute for Child Health | Perth, Australia |
| Jorgen N. Larsson, Ph.D. (secretary) | ALK-ABELLO | Horsholm, Denmark |
| Robert C. Aalberse, Ph.D. | University of Amsterdam | Amsterdam, The Netherlands |
| Donald Hoffman, Ph.D. | East Carolina University | Greenville, NC, U.S.A. |
| Thomas A.E. Platts-Mills, M.D. Ph.D. | University of Virginia | Charlottesville, VA, U.S.A. |
| Otto Scheiner, Ph.D. | University of Vienna | Vienna, Austria |
| Martin D. Chapman, Ph.D. | INDOOR Biotechnologies, Inc. | Charlottesville, VA, U.S.A. |
| Viswanath P. Kurup, Ph.D. | Medical College of Wisconsin | Milwaukee, WI, U.S.A. |

**Table 4** Online Allergen Databases

| Database | Locator |
|---|---|
| WHO/IUIS Allergen Nomenclature | www.allergen.org[a] |
| Structural Database of Allergenic Proteins (SDAP) | http://fermi.utmb.edu/SDAP |
| Food Allergy Research and Resource Program (Farrp) | www.allergenonline.com |
| Protall | www.ifr.bbsrc.ac.uk/protall |
| ALLERbase | www.dadamo.com/allerbase |
| Allergome | www.allergome.org |
| Central Science Laboratory (York, UK) | http://www.csl.gov.uk/allergen/ |

[a] Official Web site of the WHO/IUIS Subcommittee on Allergen Nomenclature.

suggested allergen name, or assign the allergen a name, provided that the inclusion criteria are satisfied. The name will later be confirmed at a full meeting of the subcommittee. Occasionally, the subcommittee has to resolve differences between investigators who may be using different names for the same allergen, or disputes concerning the chronological order of allergen identification. These issues can normally be resolved by objective evaluation of each case.

### A. Allergen Databases

The official Web site for the WHO/IUIS Sub-committee on Allergen Nomenclature, www.allergen.org, lists all allergens and isoforms that are recognized by the subcommittee and is updated on a regular basis. Over the past 5 years, several other allergen databases have been generated by academic institutions, research organizations, and industry-sponsored groups (Table 4). These sites differ in their focus and emphasis, but are useful sources of information about allergens. The Structural Database of Allergenic Proteins (SDAP) was developed at the Sealy Center for Structural Biology, University of Texas Medical Branch, and provides detailed structural data on allergens in the WHO/IUIS nomenclature, including sequence information, PDB files, and programs to analyze IgE epitopes. Amino acid and nucleotide sequence information is also compiled in the SWISS-PROT and NCBI databases. The Farrp and Protall databases focus on food allergens and provide sequence similarity searches (Farrp) and clinical data (skin tests, provocation tests) (Protall). The Allergome database provides regular updates on allergens from publications in the scientific literature. The reader is referred to Table 4 to ascertain which of these sites may be of interest.

## V. CONCLUDING REMARKS

The three men in a boat did a remarkably good job! The use of the systematic allergen nomenclature has been extremely successful and has significantly enhanced research in the area. The current list comprises 353 allergens and 190 isoallergens. The nomenclature continues to be revised. One topic under discussion is whether it is valid to include an allergen in the system if it has been demonstrated to cause IgE-mediated reactions in only five patients (the present policy) or represents <5% of a particular patient population. The problem with including allergens according to these criteria is that the number of allergens becomes very large and, unless the allergens are used in research or clinical studies, an

element of redundancy is built into the system. Conversely, it has been argued that the nomenclature is only a standardized name that permits precise communication about a particular allergen and that relative allergenic influence is not necessarily significant, provided that allergenic activity is clearly documented.

Another topic that continues to evoke discussion is the use of the generic terms "major" and "minor" in reference to an allergen. Relatively few allergens fulfill the criteria originally used by Marsh to define a major allergen (i.e., one that causes IgE response in ≥90% of allergic patients, such as Bet v 1, Fel d 1, Der p 2, or Lol p 1). However, there are a large number of allergens that cause sensitization in >50% of patients, and Lowenstein used this figure (50%) to define major allergens in the early 1980s (6). Scientists like to describe their allergens as "major" because this is effective in promoting their research and carries some weight in securing research funding. The question continues to be, "What defines a major allergen?" Demonstrating a high prevalence of IgE-mediated sensitization and that the protein has allergenic activity in vivo is a minimal requirement, given the increasing sensitivity of assays to detect IgE antibodies. The contribution of the allergen to the total potency of the vaccine should be considered (e.g., by absorption studies), as well as the amount of IgE antibody directed against the allergen, compared with other allergens purified or cloned from the same source. Other criteria include whether the allergen induces strong T-cell response and, for indoor allergens, whether it is a suitable marker of exposure in house dust and air samples. All of these criteria need to be taken into account, and ultimately, the onus is on researchers to establish the importance of their allergens by designing more creative and objective experiments.

For most purposes, allergists need only be familiar with the nomenclature for allergens (Lol p 1, Amb a 1, etc.), rather than isoallergens and peptides, for example. As measurements of allergens in extracts/vaccines or for environmental exposure become a routine part of the care of allergic patients, allergists will need to know what the allergens are and how to distinguish them. Having a systematic nomenclature will help this process. However, the nomenclature of isoallergens and variants will largely be used by researchers, allergen manufacturers, and biotechnology companies that need to identify minor differences between allergens. The systematic nomenclature is a proven success and is versatile enough to evolve with advances in molecular biology and protein science that will occur over the next decade.

## VI. SALIENT POINTS

1. A systematic nomenclature for all allergens that cause disease in humans has been formulated by a subcommittee of the World Health Organization and the International Union of Immunological Societies.
2. Allergens are described using the first three letters of the genus, followed by a single letter for the species and an arabic numeral to indicate the chronological order of allergen purification (for example, *Dermatophagoides pteronyssinus* allergen 1 = Der p 1).
3. To be included in the systematic nomenclature, allergens have to satisfy criteria of biochemical purity and criteria to establish their allergenic importance. It is important that the molecular structure of an allergen is defined without ambiguity and that allergenic activity is demonstrated in a large, unselected population of allergic patients.

4. Modifications of the nomenclature are used to identify isoallergens, isoforms, allergen genes, recombinant allergens, and synthetic peptides. For example, Bet v 1.10 is an isoallergen of Bet v 1, and Bet v 1.0101 is an isoform or variant of the Bet v 1.10 isoallergen. The prefixes "r" and "s" denote recombinant and synthetic peptides of allergens, respectively. Allergen genes are denoted by italics; e.g., *Fel d* 1A and *Fel d* 1B are the genes encoding chain 1 and chain 2 of Fel d 1, respectively.

This chapter has reviewed the systematic IUIS allergen nomenclature as revised in 1994. Other views expressed in the chapter are personal opinions and do not necessarily reflect the views of the IUIS Subcommittee on Allergen Nomenclature. The nomenclature is being updated, and a third revision is expected to be published by 2004. The author is grateful to Drs. Anna Pomés and Jorgen Larsen for assistance in preparing this chapter.

## REFERENCES

1. King TP, Norman PS. Isolation of allergens from ragweed pollen. Biochemistry 1962; 1:709–720.
2. Johnson P, Marsh DG. The isolation and characterization of allergens from the pollen of rye grass (*Lolium perenne*). Eur Polymer J 1965; 1:63–77.
3. Ohman JL, Lowell F, Bloch KJ. Allergens of mammalian origin: III. Properties of major feline allergen. J Immunol 1974; 113:1668–1677.
4. Elsayed S, Aas K. Characterization of a major allergen (cod): Chemical composition and immunological properties. Int Arch Allergy Appl Immunol 1970; 38:536–548.
5. Marsh DG. Allergens and the genetics of allergy. In: The Antigens, vol III. (Sela M, ed.) New York: Academic Press, 1975: 271–350.
6. Lowenstein H. Quantitative immunoelectrophoretic methods as a tool for the identification and analysis of allergens. Prog Allergy 1978; 25:1–62.
7. DeWeck A, Ring J. Collegicum Internationale Allergologicum: History and Aims of a Special International Community Devoted to Allergy Research, 1954–1996. Munich: MMV Medizin Verlag.
8. Marsh DG, Goodfriend L, King TP, Lowenstein H, Platts-Mills TAE. Allergen nomenclature. Bull World Health Organ 1986; 64:767–770.
9. Scheiner O, Kraft DG. Basic and practical aspects of recombinant allergens. Allergy 1995; 50:384–391.
10. Thomas WR. Molecular analysis of house dust mite allergens. In: Allergic Mechanisms and Immunotherapaeutic Strategies (Roberts AM, Walker MR, eds.). Chichester: John Wiley & Sons, 1997; 77–98.
11. Platts–Mills TAE, Vervloet D, Thomas WR, Aalberse RC, Chapman MD. Indoor allergens and asthma: Report of the third international workshop. J Allergy Clin Immunol 1997; 101:S1–S24.
12. King TP, Hoffman, Lowenstein H, Marsh DG, Platts-Mills TAE, Thomas WR. Allergen nomenclature. Bull World Health Organ 1994; 72:797–80.
13. King TP, Hoffman D, Lowenstein H, Marsh DG, Platts–Mills TAE, Thomas WR. Allergen nomenclature. Int Arch Allergy Appl Immunol 1994; 105:224–233.
14. King TP, Hoffman D, Lowenstein H, Marsh DG, Platts–Mills TAE, Thomas W. Allergen nomenclature. Allergy 1995, 50(9):765–774.
15. Chapmpan MD, Smith AM, Vailes LD, Arruda K, Dhanaraj V. Recombinant allergens for diagnosis and therapy of allergic diseases. J Allergy Clin Immunol 2000;106:409–418.
16. Pomés A, Smith AM, Grégoire C, Vailes LD, Arruda LK, Chapman MD. Functional properties

of cloned allergens from dust mite, cockroach, and cat—Are they relevant to allergenicity? ACI Int 2001; 13:162–169.

17. Ichikawa K, Iwasaki E, Baba M, Chapman MD. High prevalence of sensitization to cat allergen among Japanese children with asthma, living without cats. Clin Exp Allergy 1998; 29:754–761.
18. Trombone APF, Tobias KRC, Ferriani VPL, Schuurman J, Aalberse RC, Smith AM, Chapman MD, Arruda LK. Use of chimeric ELISA to investigate immunoglobulin E antibody responses to Der p 1 and Der p 2 in mite-allergic patients with asthma, wheezing and/or rhinitis. Clin Exp Allergy 2002; 32:1323–1328.
19. Vailes LD, Sun AW, Ichikawa K, Wu Z, Sulahian TH, Chapman MD, Guyre PM. High-level expression of immunoreactive recombinant cat allergen (Fel d 1): Targeting to antigen-presenting cells. J Allergy Clin Immunol 2002; 110:757–762.
20. Smith AM, Benjamin DC, Hozic N, Derewenda U, Smith WA, Thomas WR, Gafvelin G, van Hage-Hamsten M, Chapman MD. The molecular basis of antigenic cross-reactivity between the group 2 mite allergens. J Allergy Clin Immunol 2001; 107:977–984.
21. Gafvelin G, Johansson E, Lundin A, Smith AM, Chapman MD, Benjamin DC, Derewenda U, van Hage-Hamsten M. Cross-reactivity studies of a new group 2 allergen from the dust mite *Glycyphagus domesticus*, Gly d 2, and group 2 allergens from *Dermatophagoides pteronyssinus, Lepidoglyphus destructor*, and *Tyrophagus putrescentiae* with recombinant allergens. J Allergy Clin Immunol 2001; 107(3):511–518.
22. Derewenda U, Li J, Derewenda Z, Dauter Z, Mueller GA, Rule GS, Benjamin DC. The crystal structure of a major dust mite allergen Der p 2 and its biological implications. J Mol Biol 2002; 318:189–197.
23. Melen E, Pomes A, Vailes LD, Arruda KL, Chapman MD. Molecular cloning of Per a 1 and definition of the cross-reactive Group 1 cockroach allergens. J Allergy Clin Immunol 1999; 103:859–864.
24. Griffith IJ, Craig S, Pollock J, Yu XB, Morganstern JP, Rogers BL. Expression and genomic structure of the genes encoding Fd 1, the major allergen from the domestic cat. Gene 1992; 113:263–268.