

Arbeiten aus dem Paul-Ehrlich-Institut
11th International Paul-Ehrlich-Seminar, Bad Homburg 2005
Verlag Chmielorz

Allergen Nomenclature: A need for the scientific community

Jørgen N. Larsen

Historical introduction

Historically, the first allergen was described by King and Norman in 1962 [1]. The allergen was the major pollen allergen of *Ambrosia artemisiifolia*, short ragweed, and the authors named it Antigen E. Later Marsh isolated an important allergen from rye grass, *Lolium perenne*, and named it Rye 1 to indicate that it was the first allergen characterized from this species [2]. In the sixties and seventies several allergens from ragweed, rye grass, insect venoms and others were purified by biochemical methodology, and characterized using human sera and antibodies raised by immunization of rabbits. With the introduction of the crossed immunoelectrophoretic techniques by Løwenstein in the seventies there was a drastic increase in the number of allergens identified [3]. In most cases the biological function of the identified allergens was unknown, and names assigned were arbitrary or derived from the technique used for identification. In these years the allergen research expanded, several groups were active, and the need for a unified nomenclature became obvious.

The origins of the systematic allergen nomenclature can be traced to a meeting between David Marsh, Henning Løwenstein, and Thomas Platts-Mills on a boat ride on lake Bodensee, Germany, during the 13th Symposium of the Collegium Internationale Allergologicum (CIA) in 1980 [4]. Later an Allergen Nomenclature Subcommittee was formed under the auspices of the World Health Organization (W.H.O.) and the International Union of Immunological Societies (I.U.I.S.), and criteria for including allergens in the systematic nomenclature were established [5]. These included strict criteria for biochemical purity, as well as criteria for determining the allergenic activity of the purified protein. Furthermore, a procedure for investigators to submit data on newly defined allergens for inclusion in the systematic nomenclature was established. The chosen basis of the Allergen Nomenclature was the Linnéan taxonomy with numerals added to indicate individual allergens in order of their identification. It was decided to describe the allergen using the first three letters of the genus and the first letter of the species in italics, followed by a

roman numeral. Thus, the ragweed antigen E became *Amb a I*, and Rye 1 became *Lol p I*. The Allergen Nomenclature Sub-committee prepared an official list of allergens fulfilling the criteria, and the list comprising 27 allergens was published with the Allergen Nomenclature in the Bulletin of the World Health Organization in 1986 [5].

The systematic Allergen Nomenclature was immediately accepted by allergy researchers and clinicians. It proved to be clear, consistent and comprehensible. The new allergen names were used at meetings and in the literature, and the official list of approved names expanded rapidly to include newly identified allergens.

The current allergen nomenclature

In the late eighties the advancement of molecular biology led to new ways of identifying allergens, and advances in information technology enabled new ways of analyzing molecular data. In a single cloning experiment several IgE binding polypeptides could be identified and homology searches in the sequence databases, accessible via the Internet, could immediately establish the biological function by sequence homology. In addition, gene cloning experiments helped to establish that most, if not all, allergens are heterogeneous. The implication of 'isoallergenic variation' is that similar molecules exist with amino acid substitutions; up to 25% of the amino acids may be substituted corresponding to the difference between homologous allergens from related species, and with respect to nomenclature, each isoallergen should be identifiable by a separate name. Taken together, the technical development demanded a revision of the Allergen Nomenclature.

The revised allergen nomenclature abolishes the use of italics and roman numerals. Allergen names are now in plain writing and Arabic numerals. New in the revised allergen nomenclature are definitions of names concerning isoallergens and variants, allergen encoding genes, mRNAs and cDNAs, as well as allergenic peptides, whether of recombinant or synthetic origin, and whether modified or not. Most importantly, however, according to the revised nomenclature biochemical and serological data is not any longer considered sufficient data for unambiguous identification, sequence information is mandatory. The revised nomenclature was published in the Bulletin of the World Health Organization in 1994 [6], and also appeared in several allergy journals [7-10]. A summary of the revised allergen nomenclature is given below.

Major and minor allergens

The revised allergen nomenclature designates an allergen as major or minor depending on whether greater or less than 50% of patients allergic to the relevant allergen source have the corresponding allergen-specific IgEs.

Allergen names

Allergens are designated according to the accepted taxonomic name of their source as follows: the first three letters of the genus, space, the first letter of the species, space, and an Arabic number. The numbers are assigned to the allergens in the order of their identification, and the same number is generally used to designate homologous allergens of related species. In some instances the system needs to be

modified and an additional letter has to be added for designation of the exact genus or species. For example an 'i' is added to names of allergens of *Vespula vidua* (Ves vi) in order to distinguish them from allergens from *Vespula vulgaris* (Ves v).

Allergens and isoallergens

An allergen from a single species may consist of several highly similar molecules. These similar molecules are designated as isoallergens when they share the following common biochemical properties: 1) similar molecular weight; 2) identical biological function, if known, e.g. enzymatic action; and 3) > 67% identity of amino acid sequences. The 67% identity limit is arbitrary, and there are likely to be borderline cases.

Homologous allergens from related species belong to the same group and should have identical numbers. Homologous allergens from related species often share the above common biochemical properties, however, need not satisfy the 67% sequence identity requirement. As a rule of thumb 'related' is defined as within taxonomical order, however, it is realized that taxonomical orders are very heterogeneous with respect to number and similarity of species included. It is therefore not meaningful to apply this rule in all cases.

Isoallergens and variants

Members of an allergen group within the same species are designated as isoallergens, according to the abovementioned criteria. Each isoallergen may have multiple forms of closely similar sequences, which are designated as variants. Isoallergens and their variants are designated by suffixes of a period followed by four Arabic numerals. The first two numerals (01 to 99) refer to a particular isoallergen, and the two subsequent numerals (01 to 99) refer to a particular variant of a particular isoallergen. In many cases it is unnecessary to specify the isoallergen or variant and the corresponding suffixes may be omitted. For example Amb a 1 represents any Amb a 1 allergen and Amb a 1.0101 represents variant number 1 of isoallergen Amb a 1.01.

Allergen encoding genes, mRNAs and cDNAs

Allergen encoding genes are designated using italicized characters. Normal characters are used for designation of mRNAs and cDNAs. For example, Fel d 1 is a protein consisting of two polypeptide chains, which are encoded by two separate genes. The two allergen genes for chains 1 and 2 of Fel d 1 will be represented by *Fel d 1A* and *Fel d 1B*, respectively. Allelic forms of mRNAs and cDNAs of the *Fel d 1A* gene are designated as mRNA or cDNA Fel d 1A.0101 where the numerals are to correspond to those of the polymorphic allergens.

Recombinant allergens and peptides of allergenic interest

The nomenclature for recombinant and synthetic peptides of allergenic interest is based on the nomenclature for naturally occurring allergens. An allergen which is prepared by recombinant (r) or chemical synthetic (s) means is to be differentiated from a natural (n) allergen by the addition of the prefix of r or s followed by a suffix of the amino acid residue positions which are given in parenthesis. For examples, a recombinant house dust mite allergen Der p 1.0101 which contains the entire

sequence of 222 residues will be designated as rDer p 1.0101, and a recombinant or synthetic peptide of residue 110-131 of Der p 1.0101 will be designated as r or sDer p 1.0101 (110-131), respectively.

The revised allergen nomenclature further describes rules for the nomenclature of modified synthetic or recombinant peptides. This nomenclature is very similar to that used for synthetic peptides derived from immunoglobulin sequences [11]. Readers with special interest are referred to the Allergen Nomenclature for details [9].

Managing of the Allergen Nomenclature

The I.U.I.S. Allergen Nomenclature Sub-committee maintains the official 'list of allergens', a database of selected characteristics of each designed allergen, isoallergen and variant. The list is available on the Internet at the address <http://www.allergen.org>, the website of the Sub-committee. The website furthermore contains information on committee members, relevant publications, and a 'new allergen form' intended to be used for proposing new allergen names to be included in the list. The procedure is simple. Having characterized a new allergen and demonstrated its allergenicity, investigators should submit the completed form to the sub-committee prior to publishing articles describing the allergen. The committee will resolve any issues pertaining to the proposed name, if necessary after consulting other investigators operative in the particular area, to ensure common acceptance of the new name. The investigator will then receive a designation that can be used in the first published article describing the allergen, and the allergen will be added to the official 'list of allergens'.

The I.U.I.S. Allergen Nomenclature Sub-committee meets annually at international allergy meetings and reviews proposals received during the year, as well as other matters regarding procedures and policies. The meetings are open to the public, as is the committee-at-large, members of which receive decisions made by the sub-committee.

A fundamental issue currently under discussion in the subcommittee is the criteria for inclusion of allergens in the list. One viewpoint questions the appropriateness of including an allergen, which has only been demonstrated to bind IgE in five sera or represent a positive reaction in less than 5% of a patient population (the present policy). The problem is that the list will continue to grow comprising an increasing number of allergens of limited interest. Another viewpoint is that the nomenclature is merely a standardized system securing unambiguous communication about a particular allergen, and that relative importance is not the principal thing as long as allergenicity is clearly documented.

A compromise could be to rate the allergens in the list according to the proportion of documentation provided in the literature in terms of number of patients studied and prevalence of reactivity. Difficulties in this respect, however, are geographical variation in the relative importance of an allergen, and that relative importance is a matter of quantitative contribution to the total IgE specific for the source, which is currently difficult to measure, and not only a question of prevalence of reactivity.

These considerations lead to the principal discussion concerning criteria for the establishment of relative importance. Most scientists like to describe their allergens as 'major' in order to attract admiration and efficiently secure funding for their research. As discussed above relative importance is not just a matter of prevalence of reactivity, quantitative aspects also needs to be considered, and possibly also T cell reactivity. After all there may not be that many really important allergens, and the criteria for the designation of a major allergen, a prevalence of 50% in a patient population sensitized to the source, may be too wide. On the other hand, allergic patients suffering from allergy towards inferior allergens have an equal claim for information and care.

The initiative behind the allergen nomenclature has so far been very successful. The allergen nomenclature has been widely accepted. By the end of September 2005 the list comprises exactly 500 allergens and 275 isoallergens, and the list is continually being revised. The list serves several purposes. It facilitates knowledge dissemination and research in the well established clinically important major allergens, and it secures unambiguous communication among researchers, allergen producers, and authorities concerning the particulars of isoallergens, variants, genes, and epitopes.

References

1. King TP, Norman PS. Isolation studies 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 Polym J* 1965; 1: 63-77.
3. Løwenstein H. Quantitative immunoelectrophoretic methods as a tool for the analysis and isolation of allergens. *Prog Allergy* 1978; 25: 1-62.
4. Chapman MD. Allergen nomenclature. *Clin Allergy Immunol* 2004; 18: 51-64.
5. Marsh DG, Goodfriend L, King TP, Løwenstein H, Platts-Mills TAE. Allergen nomenclature. *Bull World Health Organ* 1986; 64: 767-774.
6. WHO-IUIS Allergen Nomenclature Subcommittee. Allergen nomenclature. *Bull World Health Organ* 1994; 72: 797-806.
7. King TP, Hoffman D, Løwenstein H, Marsh DG, Platts-Mills TAE, Thomas W. Allergen nomenclature. WHO/IUIS Allergen Nomenclature Subcommittee. *Int Arch Allergy Immunol* 1994; 105: 224-233.
8. WHO-IUIS Allergen Nomenclature Subcommittee. Allergen nomenclature. *Clin Exp Allergy* 1995; 25: 27-37.
9. World Health Organization/IUIS Allergen Nomenclature Subcommittee, Hoffman D, Løwenstein H, Marsh DG, Platts-Mills TAE, Thomas W. Allergen Nomenclature. *J Allergy Clin Immunol* 1995; 96: 5-14.
10. King TP, Hoffman D, Løwenstein H, Marsh DG, Platts-Mills TAE, Thomas W. Allergen nomenclature. *Allergy* 1995; 50: 765-774.
11. WHO-IUIS Nomenclature Subcommittee. Nomenclature for synthetic peptides representative of immunoglobulin chain sequences. *Bull World Health Organ* 1990; 68: 109-114.

Discussion

Van Ree: I almost do not dare to ask, since I am a member of the standardization committee. What is the difference of an isoallergen and a variant?

Larsen: There is actually not a stringent way of answering that question. If you look at isoallergens and variants in most systems, it will appear logical. For Amb a 1 in the

ragweed system, e.g. you have these four molecules that are actually so different from each other that they can be separated by biochemical techniques and this was presumably done in the early seventies. But still you have minor differences in these isoallergen groups. But with other allergens you do not have this, so we just number all of them as variants. I mean we have a firm rule for distinguishing allergens from isoallergens that is 66% sequence identity. And that is basically because the difference between Amb a 1 and Amb a 2 was 67%, and maybe they are actually the same allergen rather than different allergens. So I do not think we can come up with a clear answer to that.

Lorenz: I remember that the definition for isoallergens is that they have 95% identity on the protein level and that variants are more than 98% identical.

Larsen: Yes, that is what you typically find.

Moingeon: Is it sufficient to have a partial sequence?

Larsen: That could definitely be sufficient. We just need enough of the sequence to be able to unambiguously identify the molecule.

Moingeon: And show that it is different from other molecules?

Larsen: Yes, but there is a problem with that. Because sometimes people do not want sequence information to be open to the public before they publish the paper. Nevertheless, we have to be pragmatically if this system is to be used. Therefore, when people promise that they have the information, and they claim that the molecule is e.g. homologous to profilin, the committee can accept that.

Ferreira: How do you reconsult those two nomenclature rules where first you name allergens in the order they appeared, with the number of a homologous allergen? For example, one can name Bet v 1 because it was the first one discovered, but then it could happen that in hazel, the analogous allergen was the fourth. So how do you reconsult these two situations?

Larsen: Ideally we should leave empty spaces in the numbering. So, if you have characterized allergens 1, 2 and 3 in one species, numbers 4, 5, 6 in another, when you characterize the first allergen in a third species within that order this should be number 7 if it is not homologous to any of the others. But that is not always the case because the committee is not always consulted, so you may have five species within the same order where five different molecules are number 1. We do not necessarily do anything about that until there is a problem. If we have some activity in that area and it becomes a problem, then we will have to rename allergens somehow to make it fit with the system. If there is no activity in the area and we do not see a problem we just leave it as it is.