

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Publications in Food Science and
Technology

Food Science and Technology Department

11-19-2019

Keeping Allergen Names Clear and Defined

Sanny K. Chan

National Jewish Health, Denver

Anna Pomés

INDOOR Biotechnologies, Inc. Charlottesville

Christiane Hilger

Luxembourg Institute of Health, Esch-sur-Alzette

Janet M. Davies

Queensland University of Technology & Metro North Hospital and Health Service, Brisbane,
j36.davies@qut.edu.au

Geoffrey Mueller

National Institute of Environmental Health Sciences, Durham

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.unl.edu/foodsciefacpub>

 Part of the [Food Science Commons](#)

Chan, Sanny K.; Pomés, Anna; Hilger, Christiane; Davies, Janet M.; Mueller, Geoffrey; Kuehn, Annette; Lopata, Andreas L.; Gadermaier, Gabriele; Hage, Marianne van; Raulf, Monika; and Goodman, Richard E., "Keeping Allergen Names Clear and Defined" (2019). *Faculty Publications in Food Science and Technology*. 326.
<https://digitalcommons.unl.edu/foodsciefacpub/326>

This Article is brought to you for free and open access by the Food Science and Technology Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications in Food Science and Technology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Sanny K. Chan, Anna Pomés, Christiane Hilger, Janet M. Davies, Geoffrey Mueller, Annette Kuehn, Andreas L. Lopata, Gabriele Gadermaier, Marianne van Hage, Monika Raulf, and Richard E. Goodman



Keeping Allergen Names Clear and Defined

Sanny K. Chan¹, Anna Pomés², Christiane Hilger³, Janet M. Davies^{4,5}, Geoffrey Mueller⁶, Annette Kuehn³, Andreas L. Lopata⁷, Gabriele Gadermaier⁸, Marianne van Hage⁹, Monika Raulf¹⁰ and Richard E. Goodman^{11*} on behalf of the WHO/IUIS Allergen Nomenclature Sub-Committee

¹ Division of Allergy and Immunology, Department of Pediatrics, National Jewish Health, Denver, CO, United States, ² INDOOR Biotechnologies, Inc. Charlottesville, VA, United States, ³ Department of Infection and Immunity, Luxembourg Institute of Health, Esch-sur-Alzette, Luxembourg, ⁴ Centre for Children's Health Research, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD, Australia, ⁵ Metro North Hospital and Health Service, Brisbane, QLD, Australia, ⁶ National Institute of Environmental Health Sciences, Durham, NC, United States, ⁷ Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, QLD, Australia, ⁸ Department of Biosciences, University of Salzburg, Salzburg, Austria, ⁹ Division of Immunology and Allergy, Department of Medicine Solna, Karolinska Institutet and University Hospital, Solna, Sweden, ¹⁰ Institute of Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr-Universität Bochum, Bochum, Germany, ¹¹ Food Allergy Research and Resource Program, Department of Food Science and Technology, University of Nebraska-Lincoln, Lincoln, OR, United States

OPEN ACCESS

Edited by:

Heiko Mühl,
Goethe University Frankfurt, Germany

Reviewed by:

Peisong Gao,
Johns Hopkins University,
United States
Eva Untersmayr,
Medical University of Vienna, Austria

*Correspondence:

Richard E. Goodman
Rgoodman2@unl.edu

Specialty section:

This article was submitted to
Inflammation,
a section of the journal
Frontiers in Immunology

Received: 29 July 2019

Accepted: 21 October 2019

Published: 19 November 2019

Citation:

Chan SK, Pomés A, Hilger C, Davies JM, Mueller G, Kuehn A, Lopata AL, Gadermaier G, van Hage M, Raulf M and Goodman RE (2019) Keeping Allergen Names Clear and Defined. *Front. Immunol.* 10:2600. doi: 10.3389/fimmu.2019.02600

The World Health Organization/International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature Sub-Committee was established in 1986 by leading allergists to standardize names given to proteins that cause IgE-mediated reactions in humans. The Sub-Committee's objective is to assign unique names to allergens based on a critical analysis of confidentially submitted biochemical and clinical data from researchers, often prior to publication to preserve consistency. The Sub-Committee maintains and revises the database as the understanding of allergens evolves. This report summarizes recent developments that led to updates in classification of cockroach group 1 and 5 allergens to animal as well as environmental and occupational allergens. Interestingly, routes, doses, and frequency of exposure often affects allergenicity as does the biochemical properties of the proteins and similarity to self and other proteins. Information required by the Sub-Committee now is more extensive than previously as technology has improved. Identification of new allergens requires identification of the amino acid sequence and physical characteristics of the protein as well as demonstration of IgE binding from subjects verified by described clinical histories, proof of the presence of the protein in relevant exposure substances, and demonstration of biological activity (skin prick tests, activation of basophils, or mast cells). Names are assigned based on taxonomy with the abbreviation of genus and species and assignment of a number, which reflects the priority of discovery, but more often now, the relationships with homologous proteins in related species.

Keywords: allergen nomenclature, WHO/IUIS, taxonomy, diagnostic, airway, food, dermal

INTRODUCTION

Advances in molecular biology, recombinant protein technology, methods of genomic, and proteomic analysis, structural biology, and high throughput screening have led to a greater understanding of allergenic proteins over the past 30 years. Identification of new allergens and improved characterization dictate the need for revision of some allergens and updated requirements for allergen nomenclature.

In the 1980's key clinical and experimental allergists led by D Marsh at Johns Hopkins University, USA, L Goodfriend (McGill University, Canada), TP King (Rockefeller University, USA), H Lowenstein (University of Copenhagen (Denmark)) and TAE Platts-Mills (University of Virginia, USA) framed basic rules for naming allergenic proteins in common allergenic sources (1). Over 2 years they agreed to a strategy of naming allergens based on taxonomy with identification using the first three letters of the genus, the first letter of the species and a number for order of identification. The first allergens were mostly inhalant proteins from pollens of weeds, trees and grasses and a few from animal dander, fungi, and venom proteins from stinging insects. The committee recognized that a systemic standardized nomenclature was needed for consistent identification in scientific publications. The standardized naming of allergens has become a formalized process where allergen names are assigned by the WHO/IUIS Allergen Nomenclature Sub-Committee after a defined submission process that includes data on IgE binding to identified novel target proteins or glycoproteins with the most recent major revision occurring in 2018 (2).

Allergen names are assigned with the first 3 or sometimes 4 letters of the genus, one or sometimes two letters for the species followed by an Arabic numeral, based on order of discovery (2). However, the same Arabic number is often used for conserved protein families in related taxa. Often species have evolved by duplication of genes with mutations including additions or deletion of nucleotides to produce alternative proteins of similar function, resulting in the occurrence of isoallergens as well as isoforms or variants within the individual organism or in related organisms of the same species. The isoallergens are designated by the addition of two digits after the decimal in the number and isoforms or variants by addition of two more digits (e.g., Amb a 1.0101). Since the evolutionary steps involved in generating each new isoform or variant is usually not known, the committee bases designations on the percent identity of amino acids compared to the first identified sequence. Sequences within ~67% identity to the original allergen are designated as isoallergens and sequences differing by <90% identity are isoform or variants.

Expression of recombinant proteins for IgE tests and characterization has led to more accurate identification. Changes to the nomenclature have attempted to incorporate an increased understanding of allergens. Higher resolution studies and isolation techniques of previously defined allergens now shows some to be composed of multiple subunits from different genes; Fel d 1, or multimers; collagen, vicilins, and glycinins. This manuscript focuses on recent challenges and underlying reasons

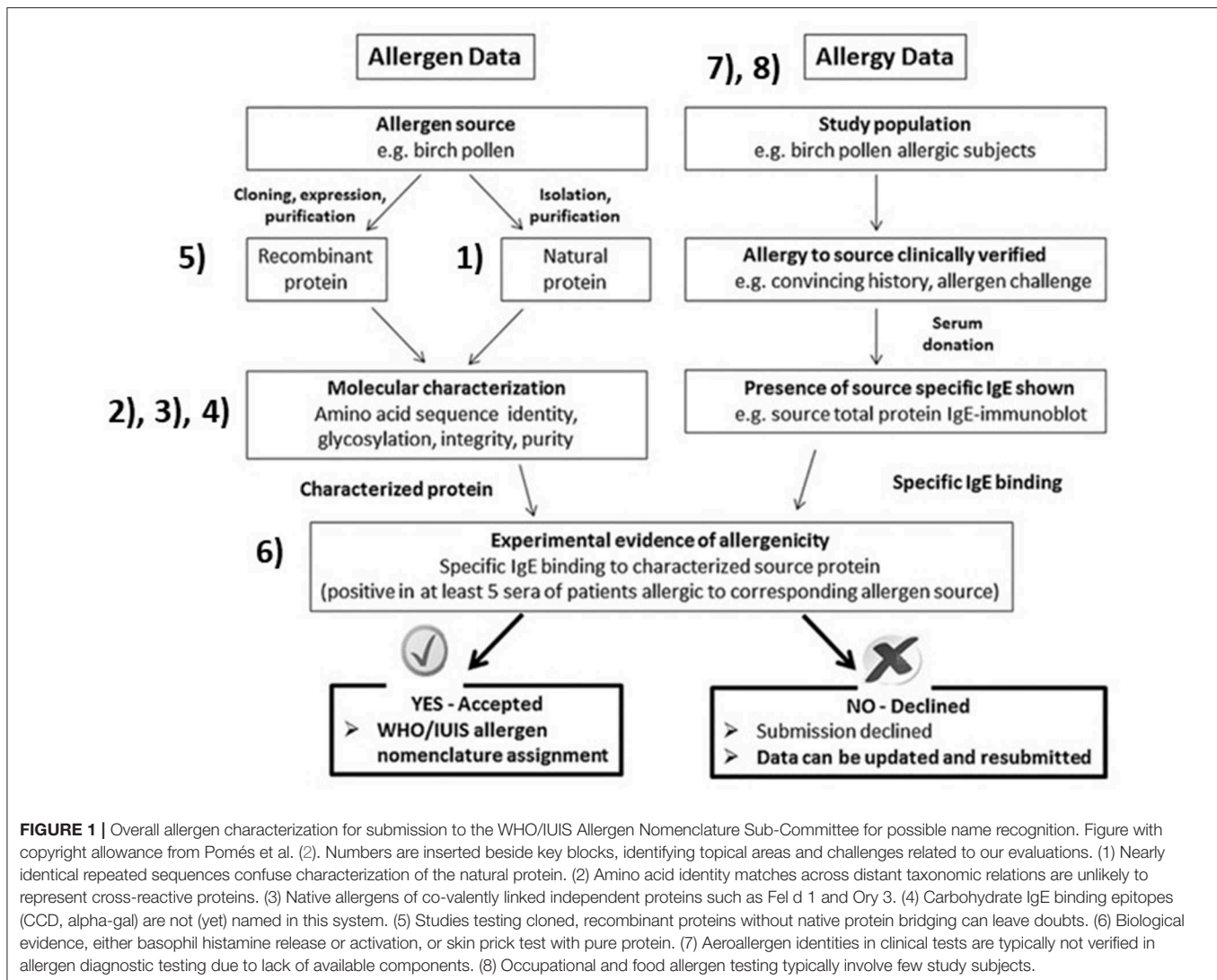
for name assignment by the WHO/IUIS Nomenclature Sub-Committee. We refer those working on allergen identification to **Figure 1**, from our 2018 publication with references to specific sections in this paper to explain the ideal process (2). The science of allergen determination has improved since 1990, for characterization of allergenic sources, protein sequences, post-translational modification determination, improved methods for measuring IgE binding, mediator release and measurements of specific bioactivity (skin tests and basophil assays). Scientists striving to characterize allergens should be aware of and using techniques appropriate to the types of allergens including those of foods, contact allergens, aeroallergens and venoms or salivary sources of allergy.

CHALLENGES FOR FUTURE ALLERGEN NAMES

Repeated Structure of Group 1 Cockroach Allergens Is Not Reflected in Current Allergen Nomenclature

Cockroach allergens from group 1 were difficult to name using the current WHO/IUIS paradigm. Cloning of the DNA encoding for Bla g 1 allergen revealed the existence of multiple DNA repeats encoding approximately 100 amino acids each (3). The molecular structure of Bla g 1 showed that two tandem amino acid repeats form a spherical fold (4). A recent analysis of the *Blattella germanica* genome (5) suggests there are 5 separate gene products, and each produce 1 to 5, partial or full repeats of the Bla g 1 structural motif. In total, there may be 9 other Bla g 1-like allergens, with 60–90% identity to the Bla g 1 structure. Prior to the genomic data, it was unclear whether the different Bla g 1 polymorphisms derived from a different gene. In addition, there was not a precedent for an allergen with this kind of structure and genomic origin.

Besides cockroaches, it is worthwhile to define new rules for assigning a nomenclature to this type of allergen, given the possibility that Bla g 1-like homologs exist in other sources, such as locusts and mealworms, recently included as food sources. Allergens belong to a selected group of protein families (6, 7), and it is likely that any Bla g 1-like homologs in these sources could also be an allergen. Indeed, a homolog of Bla g 1 was noted in the genome of *Locusta migratoria* (8). Further examination reveals that there are two genes with 2 and 3 copies of the Bla g 1-like structure, respectively. Other coleopteran species have been proposed as a substitute protein source in place of mammalian meat. A search of genomes of related species revealed that *Tribolium castaneum*, has one gene with 9 repeats similar to the Bla g 1- structure. This repeated sequence structure is likely to occur in other beetles and there are reported occupational and food allergies reported from a larval beetle (mealworm) of *Tenebrio molitor* (9). Homologs of Bla g 1 can be found in many other insect species as well (10). So far, the protein(s) responsible for reactions to mealworm are not yet identified except for tropomyosin with homology to shrimp proteins. The structural complexity of allergens like Bla g 1 are not reflected in the current allergen nomenclature. For similar cases, the



Allergen Nomenclature Sub-Committee discussed the possibility (not yet approved) of using capital letters behind the allergen name to signify either the different genes from which the allergen originated, or different duplexes derived from a gene (i.e., Bla g 1.0101A). Knowledge of the origin of unique allergen structures may help to define a more informative allergen nomenclature for these unusual allergens.

Group 5 Cockroach Allergens Are Glutathione S-Transferases (GSTs), Consideration of the Degree of Amino Acid Identity

In recent years the WHO/IUIS Sub-Committee has used similar allergen numbers for homologous proteins across related taxonomic groups when possible. Glutathione S-transferases (GST) are common enzymes present in multiple organisms from different taxonomic groups. One of the main functions of GSTs is detoxification by catalyzing the addition of a glutathione

(GSH) molecules to other commonly toxic compounds that are ultimately removed from the cell (11, 12). Nine groups of allergens are recognized as GSTs and are listed in the WHO/IUIS Allergen Nomenclature database (www.allergen.org), including those from cockroaches (Bla g 5 and Per a 5), mites (Der p 8, Der f 8 and Blo t 8), helminths (Asc s 13, Asc l 13), fungi (Alt a 13), and plants (Bet v 8) (**Figure 2A**).

In general, homologous allergens from the same species are included in the same allergen group as isoallergens (e.g., Cyn d 1.0101 and Cyn d 1.0201 in **Table S1**) if they share a suggested threshold of 67% amino acid identity (2). At 67% identity, IgE cross-reactivity is possible. However, naming additional GSTs from cockroach in group 5, has been challenging due to their low sequence identities (down to 18% identity). For those identified as binding IgE, they could be given a different name and number (**Figure 2B**). We considered these enzymes belong to different protein classes (sigma, delta, theta, mu) as functionally related. The Bla g 5 from German cockroach (*Blattella germanica*) was the first cockroach GST to be identified (O18598), and it is a

A GST allergens:

	1	2	3	4	5	6	7	8	9
1: Alt a 13.0101-AY514673-Arora	100.00	10.93	16.40	16.40	17.71	16.76	16.24	16.13	16.27
2: Bet v 8.0101-KF246508-Bohle	10.93	100.00	20.21	20.21	18.64	17.58	14.00	13.71	12.85
3: Per a 5.0101-Delta-MG255130- Tungtrongchitr	16.40	20.21	100.00	99.54	20.99	18.78	21.74	16.67	20.44
4: Per a 5.0102-Delta-AEV23867-Chew	16.40	20.21	99.54	100.00	20.99	18.78	21.74	16.67	20.44
5: Asc s 13.0101-Sigma-X75502-Leibau	17.71	18.64	20.99	20.99	100.00	37.19	23.15	26.15	25.14
6: Bla g 5.0101-Sigma-O18598-Chapman	16.76	17.58	18.78	18.78	37.19	100.00	24.24	29.17	28.49
7: Blo t 8.0101-Mu-GQ398117-Caraballo	16.24	14.00	21.74	21.74	23.15	24.24	100.00	35.48	34.36
8: Der p 8.0101-Mu-S75286-Baldo	16.13	13.71	16.67	16.67	26.15	29.17	35.48	100.00	70.56
9: Der f 8.0101-Mu-KC305499-An	16.27	12.85	20.44	20.44	25.14	28.49	34.36	70.56	100.00

B Cockroach GSTs:

	1	2	3	4	5	6	7
1: Sigma-German CR-Bla g 5.0101-O18598-Chapman	100.00	99.50	18.89	20.11	20.11	18.99	18.99
2: Sigma-German CR-EF202178-Yong	99.50	100.00	18.75	20.57	20.57	19.43	19.43
3: Theta-German CR-AEV23882-Chew	18.89	18.75	100.00	57.21	57.67	58.14	58.60
4: Delta-American CR-Per a 5.0101-MG255130- Tungtrongchitr	20.11	20.57	57.21	100.00	99.54	81.48	81.94
5: Delta-American CR-Per a 5.0102-Delta-AEV23867-Chew	20.11	20.57	57.67	99.54	100.00	81.94	82.41
6: Delta-German CR-ABX57814-Yong	18.99	19.43	58.14	81.48	81.94	100.00	98.61
7: Delta-German CR-AEV23880-Chew	18.99	19.43	58.60	81.94	82.41	98.61	100.00

Background color code according to percentage of amino acid identities:

Black: 100%, or Dark grey: Medium grey >70%: Light grey > 30% Very light grey: > 20 %: White: < 20 %.

FIGURE 2 | Homologous protein GST allergens are currently categorized as isoallergens with a threshold of 67% amino acid identity although they may belong to different protein classes (sigma, delta, theta, mu). Percent identities are shown. **(A)** Nine groups of GST allergens from different taxa to understand sequence identities across sources. **(B)** Cockroach GSTs within German and American cockroaches with known allergens and genomic sequences with comparison of amino acid sequence identities.

sigma class GSTs (11). Additional GSTs from German cockroach have been identified, though these have not been submitted to the WHO/IUIS committee with proof of IgE binding: another sigma GST (EF202178), a theta GST (AEV23882) and two delta GSTs (ABX57814 and AEV23880). Two GSTs from the American cockroach (*Periplaneta americana*) were identified as binding IgE, and they belong to the delta class of GSTs [MG255130 (13) and AEV23867] and share 99.54% amino acid identity. When these GSTs were submitted, we decided to include them as Per a group 5 proteins even though their identity to Bla g 5 was only 19% identity with Bla g 5. They were named Per a 5.0101 and Per a 5.0102. If other cockroach GSTs are submitted to the database in the future, they will be assigned as isoallergens if fitting within the 67% identity range to the respective group 5 allergens.

GSTs from other species are and have been assigned to other numbers as seen in **Figure 2A** (Alt a 13, Bet v 8, Asc a 13, Blo t 8, Der p 8 and Der f 8). Isoallergens of each of these will be assigned, based on sequence comparison, to currently named members of the respective source species. The committee now prefers to retain primary allergen numbers the same across related taxonomic groups, but that is not always possible. It is important to note that while enzyme functions may be retained with great diversity in amino acid sequences, the ability for IgE to bind across those species and meet the criteria of allergenicity is not retained.

Covalently Linked Dimeric Allergens and Allergen Nomenclature

Mammalian allergens of the secretoglobulin family are heterodimers, encoded by separate genes. Secretoglobins (SCGB) represent a large protein family found in mammals and

marsupials. Common characteristics are secretion in many body fluids, small size, alpha-helical, and dimeric structure creating a hydrophobic binding pocket (14, 15). Major allergens from cat and rabbit are Fel d 1 and Ory c 3. Both are glycosylated heterodimers linked by three disulfid bridges (16, 17). Fel d 1 also forms tetramers composed of two heterodimers. Each protein chain of the heterodimer is encoded by a single gene and both are needed to compose the natural molecule and may be needed for achieving full allergenic activity. The current allergen nomenclature system did not foresee linked heteromeric allergens and the names allocation for Fel d 1 and Ory c 3 were not consistent. Whereas, Fel d 1 is listed as Fel d 1.0101 for the dimer, composed of a chain 1 and a chain 2, products of two separate genes (17), Ory c 3 is listed as 2 isoallergens with Ory c 3.A.0101, chain A (lypophillin CL2) and Ory c 3.B.0101, chain B (lypophillin AL). The nomenclature for Ory c 3 suggests that both isoallergens have to be considered together as a full allergen (17). The two proteins do not exist naturally as monomers and allergenicity can not be assessed individually. The Sub-Committee recently renamed the Fel d 1 proteins as Fel d 1.A.0101 for chain 1 and Fel d 1.B.0101 for chain 2, consistent with the rabbit allergen.

Other examples are recently added allergens, barramundi collagen Lat c 6 and salmon allergen Sal s 6 proteins. Collagen is a triple-stranded rope-like coiled structure, composed of 3 protein chains. Each chain has been given an isoallergen name, despite the fact that they do not exist as monomers. Currently it is only possible to evaluate separate IgE binding if using recombinant monomeric proteins and thus it is not possible to assess the allergenicity of each of the entire native proteins. The committee may reevaluate this policy in the future.

Carbohydrates, New Allergen Epitopes Not Represented in the Allergen Nomenclature Database

Cross-reactive carbohydrate determinants (CCDs) are carbohydrate epitopes carrying an α -1,3-fucose and, or β 1,2 xylose on complex carbohydrates of some glycoproteins are the main targets of IgE recognition. They are present on many plant and insect proteins and responsible for extensive IgE cross-reactivity between different allergen sources (18). However, there is a general consensus that those glycans do not trigger noticeable clinical symptoms. Two other carbohydrate groups have been associated with food allergy and anaphylaxis (19). Galactose- α -1,3-galactose (α -gal), a carbohydrate present on mammalian proteins, has been reported to trigger delayed severe allergic reactions to red meat and to induce acute allergic reactions upon injection of therapeutic antibodies carrying α -gal (20, 21). Galacto-oligosaccharides (GOS) are heat-stable, non-digestible carbohydrates present in milk formula and as supplement in different beverages. Food allergy to GOS are reported in Singapore, Vietnam and Japan (19). The structure and characterization of carbohydrates that are the reported targets of IgE binding have not been demonstrated, although they have been assumed to be related to the synthesis of GOS complexes. Recent work failed to identify the structures or link them to activation of basophils (22). In contrast to α -gal, the GOS carbohydrates have not been linked to a protein backbone (19). So far, only allergens representing protein epitopes have been named by the WHO/IUIS Nomenclature Sub-Committee. The committee is currently considering whether there is enough definition of carbohydrate epitopes to provide unique names and whether names would be scientifically useful.

Updating Existing Allergen Entries Named From Partial Allergen Sequences

Some allergens are characterized based on peptide sequences from native molecules. In some cases, full-length sequences were not available, but are based on peptide matches to genomic sequences. That was the case for fish allergens, enolase Sal s 2 and aldolase Sal s 3 from salmon (*Salmon salar*) based on public genomic sequences (23). For the salmon allergens Sal s 2 and Sal s 3, complete sequences were attributed without testing of the allergenic properties of the corresponding proteins. In the future, we recommend that an alignment of the complete sequences of natural or recombinant protein be used to determine IgE binding be compared to the genome sequence, or at the least, a test of the recombinant protein be comparison to the native counterpart. Importantly, different isoforms and variants can exist in the same source and although highly similar, those proteins can differ in their allergenic potency (24). Verified information will be added to the existing database entry in order to further improve its quality.

Numbering of Allergens

The allergen nomenclature system was originally based on the order of allergen discovery to provide the allergen number, but

the committee recognized the conservation of sequences and structures are important. But sometimes, very similar allergens from related species receive different numbers (e.g., Fel d 2 and Can f 3) (25). It would be disruptive to change allergen names after a number of publications have used established allergen names and only allergen numbers have been revised and changed in a few cases (26).

OVERVIEW OF OUTDOOR ENVIRONMENTAL AEROALLERGENS

Most outdoor aeroallergens are derived from grass, weed and tree pollen or fungal spore sources. Although Blackley discovered that pollen caused hay fever in the nineteenth century (27), it was not until allergenic fractions were purified by chromatography from ryegrass pollen in the 1960s (28) and cloned in 1990s (29) that the first pollen allergens were characterized. Because the sources are typically not recognized due to small sizes of pollen (10–50 microns) and mold spores (2–50 microns), there is often uncertainty regarding the causal link between exposure to a given species and allergic reactions. This is specifically difficult given the cross-reactivities that exists among allergens from different species.

The IUIS Allergen Nomenclature Subcommittee recognize 43, 55, and 45 allergens for pollen of grasses, trees and weeds, respectively. Another four pollen allergens of herbaceous plants included; Lig v 1 of *Ligustrum vulgare* (Common privet), Syr v 1 and Syr v 3 of *Syringa vulgaris* (Lilac) and Hum j 1 of *Humulus japonicus* (Japanese hop).

Grass pollen allergens belong to 12 different protein families. The most abundant proteins of mature pollen that elicits a high frequency of IgE reactivity are the major group 1 beta-expansin allergens. The purpose of beta-expansins is to degrade cell walls and allow pollen tube extension which is essential for fertilization. A number of IgE and T cell epitope regions are found in close proximity with the enzymatic domain and specifically the HFD motif of the N terminal domain (30–32). While allergens of temperate grass pollens were characterized in 1990s, allergens of subtropical grass pollens continue to emerge. Grass pollen group 1 allergens among the Pooideae temperate grasses share between 84 and 91% amino acid identity, while subtropical grasses of Panicoideae, Chloridoideae, and Oryzoideae share between 49 and 86% identity (Table S1). Separate gene loci encode multiple Poaceae group 1 isoforms. The pollen expansin proteins show significant sequence diversity, but they apparently share similar functions exemplified by Cyn d 1 and Sor h 1 isoform diversity (Table S1). While research shows that there is IgE and T cell cross-reactivity between group 1 and group 5 grass pollen allergens across broad taxa, their heterogeneity contributes to the diversity of immune recognition (33, 34). The most recently listed (grass pollen allergen is Uro m 1 of the subtropical Panicoideae grass *Urochloa mutica* (Para grass). Group 2 and 3 grass pollen allergens share approximately 30% identity with the carboxyl domain of the beta-expansin protein family. Consequently, grass pollen group 2/3 pollen allergens are smaller in size (10–12 kDa), but their function is unknown. Sor h 2 and Ory s 2 are described

in subtropical species, but their clinical importance is less well-demonstrated. Several isoforms and multiple variants exist for Group 5 allergens (e.g., for Phl p 5) and these share between 53 and 78% identity between six species for which these allergens have been characterized (Table S2). Group 5 allergens are absent from subtropical grass pollen.

Weed pollen allergens arise from 16 species including ragweed (*Ambrosia artemisiifolia*), mugwort (*Artemisia vulgaris*), pellitory (*Parietaria judaica*), amaranth, thistle (*Salsola kali*), and sunflower (*Helianthus annuus*) and belong to 10 allergen families. Clinically important allergens of the Asteraceae family include pectate lyases (Amb a 1 and Art v 6), defensin-like proteins (Amb a 4 and Art v 1), non-specific lipid transfer proteins (Art v 3 and Par j 2), whereas allergens of the Amaranthaceae family include Ole e 1-like proteins and Che a 1, and pectin methylesterase (Sal k 1).

Trees from multiple orders; Fagales; for example, Birch (*Betula verrucosa*), Lamiales; Ash (*Fraxinus excelsior*) and Olive (*Olea europaea*), Proteales; Plane (*Platanus acerifolia*) and Cupressales; Japanese Cedar (*Cryptomeria japonica*), encompassing 21 species contribute allergens belonging to 18 protein families. Clinically important allergens include Pathogenesis Related protein-10 (Bet v 1, Ole e 1-like protein); Fra e 1, polygalacturonase; Cry j 2 and Pla a 2, and pectate lyase; Cry j 1. Profilins are ubiquitous and highly conserved in sequence and structure (e.g., Phl p 12, Amb a 8 and Bet v 2) and polcalcins (Phl p 7, Amb a 9, and Ole e 8) are pan-allergens common to grass, weed and tree pollen, though these allergens are generally less potent than PR-10 proteins.

A variety of fungi produce the 112 listed WHO/IUIS allergens from as many as 36 protein families. While fungal hyphae or spores may be used as food or drug sources, many are common sources of indoor and outdoor allergy and a small number colonize human tissues including skin and lungs. Spores of 15 Ascomycota species belonging to diverse fungal families; Aspergillaceae, Cladosporiaceae, Didymellaceae, Nectriaceae, and Pleosporaceae produce 72 airborne outdoor fungal spore allergens. There are 23 allergens listed for *Aspergillus fumigatus* including the ribotoxin (Asp f 1), peroxisomal protein (Asp f 3), metalloprotease (Asp f 5), cyclophilin (Asp f 11), vacuolar serine protease (Asp f 18), and enolase (Asp f 22), which are all major allergens. *Alternaria alternata* spores, one of commonly recognized fungal spores contains 12 listed allergens including its major allergens (Alt a 1), and, a glutathione S transferase (Alt a 13).

OVERVIEW OF OCCUPATIONAL ALLERGENS FROM FOOD SOURCES

Occupational sensitization has been observed for centuries in historical and medical writings. The first description of baker's asthma was probably by Ramazini around 1700; the first report of allergic responses to fish was in 1937 by De Besche relating to a fisherman who developed allergic symptoms when handling codfish (35). Sensitization to allergens at the workplace occurs usually via the skin or the respiratory tract. Resulting cutaneous

and respiratory diseases are a global health problem, with an estimate that 25% of all asthma cases related to work (36).

The most common workplace allergens are relatively high-molecular-weight proteins derived from cereal flour, natural rubber latex, wood dust, livestock and laboratory animals, seafood, industrial enzymes and mold (37). Animal and vegetable high-molecular-weight proteins present in the aerosol and often in aerosolized foods during food processing are the main inhalant allergen sources. This type of IgE-mediated respiratory food allergy usually does not generate any symptoms upon ingestion of the source material and is suggested to reflect a new type of allergy, termed Class 3 food allergy (37).

Over 400 occupational allergenic sources are identified and documented as sensitizers, derived from plants, animals and microbes. However, few have been characterized on the molecular level and only a limited number of purified native or recombinant allergens are available for testing (37). The current WHO/IUIS database categorizes allergens by typical route of exposure. However, in the case of occupational allergens the specific route is often not clear, and the route of sensitization may differ from the common route of elicitation and may be listed as "unknown."

A good example of the complex nature of identifying "occupational allergens" among plants is wheat (*Triticum aestivum*), where 28 allergens are registered with the WHO/IUIS, however only 13 are confirmed food allergens. The route of sensitization is categorized as "unknown" for 15, with 14 being relevant in baker's asthma and several allergens are relevant for grass-pollen cross-reactivity, including profilin (Tri a 12) and a subunit of the tetrameric heterologous α -amylase inhibitor chloroform/methanol-soluble CM17 protein (Tri a 40).

Among animal derived allergens, tropomyosin is an abundant and clear food allergen from crustacean shellfish and molluscs that is also implicated in IgE binding and possibly in triggering airway allergy from mites and some insects. *In vitro* IgE binding studies and skin tests without clear clinical histories of the patient can lead to uncertain diagnosis due to high amino acid homology across diverse taxonomic groups. Clear differentiation is often only possible if IgE inhibition studies are performed and verification of biological activity. Dose and exposure are important. The relevance of IgE binding to environmental sources (house dust mite, cockroach and other organisms) often leads to speculation about route of exposure in sensitization. Of the 35 registered tropomyosins, 22 are known food allergens, while 13 are not known to cause food allergies. Of these 10 are derived from insects or mites and the route of sensitization is most likely via inhalation. Among the food allergens 19 tropomyosin are derived from different seafood species and are relevant occupational allergens, however not identified as such (38). Some animal allergens are not commonly recognized by patients or clinicians due to limited exposure for most people to the source. Hemoglobins from the insect "harlequin fly larvae," *Chironomus thummi* bind IgE from sera of those who raise the insects for fish food (39). A few recent cases of occupational allergy have been noted for workers who raise mealworms (*Tenebrio molitor*) as a source of protein for processed human food (9, 40). Some individuals have been newly sensitized at work,

while for others it appears, they were sensitized by consumption of shrimp or other crustaceans. Highly conserved tropomyosin and a few enzymes such as arginine kinase may be responsible for cross reactivity.

Many other foods have been reported in single case reports as the cause of occupational allergies and asthma, and sensitization by inhaled allergen exposure is therefore quite likely (41). Importantly, though these proteins may be tolerated during exposure by one route or by most people, they may not be tolerated if exposure occurs differently and as such, the WHO/IUIS nomenclature committee has not asked for *in vivo* proof that new allergens illicit allergic responses to date, only *in vitro* evidence and a good clinical history of symptoms.

SUMMARY AND CONCLUSIONS

Though molecular and scientific advances have led to improved characterization of allergens and with-it increased clarity in naming and categorizing them, the heterogeneous nature of human data remains a challenge. This Sub-Committee has in the past asked for minimal data characterizing protein sequences, clinical symptoms of serum donors or test subjects, IgE binding and allergic functionality. We see great diversity in data submitted with candidate proteins. Our recent publications are intended to help researchers focus on relevant questions to improve the evidence that individual proteins are allergens, and not simply cross-reactive in IgE binding with only mild clinical consequences. Sequences of the allergens and demonstration of IgE binding from relevant symptomatic individuals are essential for defining allergens. Some researchers would like to simply submit to the Sub-Committee information about recombinant proteins based only on genomic sequences of potential allergen sources. While that activity may help predict possible allergens, our goal is to call for direct proof of allergenicity using subjects with clear clinical reactivity to source materials, and with clear IgE binding to proteins presented at representative concentrations.

REFERENCES

1. Marsh DG, Goodfriend L, King TP, Lowenstein H, Platts-Mills TA. Allergen nomenclature. *Bull World Health Organ.* (1986) 64:767–74.
2. Pomés A, Davies JM, Gadermaier G, Hilger C, Holzhauser T, Lidholm J, et al. WHO/IUIS allergen nomenclature: providing a common language. *Mol Immunol.* (2018) 100:3–13. doi: 10.1016/j.molimm.2018.03.003
3. Pomés A, Melén E, Vailes LD, Retief JD, Arruda LK, Chapman MD. Novel allergen structures with tandem amino acid repeats derived from German and American cockroach. *J Biol Chem.* (1998) 273:30801–7. doi: 10.1074/jbc.273.46.30801
4. Mueller GA, Pedersen LC, Lih FB, Glesner J, Moon AF, Chapman MD, et al. The novel structure of the cockroach allergen Bla g 1 has implications for allergenicity and exposure assessment. *J Allergy Clin Immunol.* (2013) 132:1420–6. doi: 10.1016/j.jaci.2013.06.014
5. Harrison MC, Jongepier E, Robertson HM, Arning N, Bitard-Feildel T, Chao H, et al. Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nat Ecol Evol.* (2018) 2:557–66. doi: 10.1038/s41559-017-0459-1
6. 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) 121:847–52. doi: 10.1016/j.jaci.2008.01.025
7. Radauer C, Breiteneder H. Pollen allergens are restricted to few protein families and show distinct patterns of species distribution. *J Allergy Clin Immunol.* (2006) 117:141–7. doi: 10.1016/j.jaci.2005.09.010
8. Spit J, Badisco L, Vergauwen L, Knapen D, Vanden Broeck J. Microarray-based annotation of the gut transcriptome of the migratory locust, *Locusta migratoria*. *Insect Mol Biol.* (2016) 25:745–56. doi: 10.1111/imb.12258
9. Broekman HCHP, Knulst AC, den Hartog Jager CF, van Bilsen JHM, Raymakers FML, Kruizinga AG, et al. Primary respiratory and food allergy to mealworm. *J Allergy Clin Immunol.* (2017) 140:600–3.e7. doi: 10.1016/j.jaci.2017.01.035
10. Randall TA, Perera L, London RE, Mueller GA. Genomic, RNAseq, and molecular modeling evidence suggests that the major allergen domain in insects evolved from a homodimeric origin. *Genome Biol Evol.* (2013) 5:2344–58. doi: 10.1093/gbe/evt182

The purpose of the allergen nomenclature system is to help researchers, clinicians, pharmaceutical companies, regulators, and the public clearly understand the identity of clinically important allergens for diagnosis and to help ensure compliance for improved safety.

However, our system is not perfect. Recent research shows that some specific glycans are important allergens including α -gal that is bound to tick and mammalian proteins and to glycolipids. We do not define these structures as allergens but hope to help publicize risks and educate consumers and allergists about risks.

AUTHOR CONTRIBUTIONS

Each author contributed a significant portion of writing for the manuscript and reviewed the final document. This collection of authors is a representative segment of the WHO/IUIS Allergen Nomenclature Sub-Committee. AK is not a Sub-Committee member, but has contributed to identification and characterization of allergens in this manuscript.

FUNDING

This work was supported by American Academy of Allergy Asthma and Immunology (AAAAI), European Academy of Allergy and Clinical Immunology (EAACI), and International Union of Immunological Societies (IUIS).

ACKNOWLEDGMENTS

This article was prepared on behalf of the WHO/IUIS Allergen Nomenclature Sub-Committee shown at www.allergen.org.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.02600/full#supplementary-material>

11. Arruda LK, Vailes LD, Platts-Mills TA, Hayden ML, Chapman MD. Induction of IgE antibody responses by glutathione S-transferase from the German cockroach (*Blattella germanica*). *J Biol Chem.* (1997) 272:20907–12. doi: 10.1074/jbc.272.33.20907
12. Mueller GA, Pedersen GL, Glesner J, Edwards LL, Zakzuk J, London RE, et al. Analysis of glutathione S-transferase allergen cross-reactivity in a North American population: relevance for molecular diagnosis. *J Allergy Clin Immunol.* (2015) 136:1369–77. doi: 10.1016/j.jaci.2015.03.015
13. Sookrung N, Reamtong O, Poolphol R, Indrawattana N, Seesua W, Saelim N, et al. Glutathione S-transferase (GST) of American cockroach, *Periplaneta americana*: classes, isoforms, and allergenicity. *Sci Rep.* (2018) 8:484. doi: 10.1038/s41598-017-18759-z
14. Klug J, Beier HM, Bernard A, Chilton BS, Fleming TP, Lehrer RI, et al. Uteroglobulin/Clara cell 10-kDa family of proteins: nomenclature committee report. *Ann N Y Acad Sci.* (2000) 923:348–54. doi: 10.1111/j.1749-6632.2000.tb05549.x
15. Jackson BC, Thompson DC, Wright MW, McAndrews M, Bernard A, Nebert DW, et al. Update of the human secretoglobin (SCGB) gene superfamily and an example of “evolutionary bloom” of androgen-binding protein genes within the mouse Scgb gene superfamily. *Hum Genomics.* (2011) 5:691–702. doi: 10.1186/1479-7364-5-6-691
16. Kaiser L, Velickovic TC, Badia-Martinez D, Adedoyin J, Thunberg S, Hallén D, et al. Structural characterization of the tetrameric form of the major cat allergen Fel d 1. *J Mol Biol.* (2007) 370:714–27. doi: 10.1016/j.jmb.2007.04.074
17. Hilger C, Kler S, Arumugam K, Revets D, Muller CP, Charpentier C, et al. Identification and isolation of a Fel d 1-like molecule as a major rabbit allergen. *J Allergy Clin Immunol.* (2014) 133:759–66. doi: 10.1016/j.jaci.2013.04.034
18. Altmann F. Coping with cross-reactive carbohydrate determinants in allergy diagnosis. *Allergo J Int.* (2016) 25:98–105. doi: 10.1007/s40629-016-0115-3
19. Soh JY, Huang CH, Lee BW. Carbohydrates as food allergens. *Asia Pac Allergy.* (2015) 5:17–24. doi: 10.5415/apallergy.2015.5.1.17
20. Hilger C, Fischer J, Wölbing F, Biedermann T. Role and mechanism of galactose- α -1,3-galactose in the elicitation of delayed anaphylactic reactions to red meat. *Curr Allergy Asthma Rep.* (2019) 19:3. doi: 10.1007/s11882-019-0835-9
21. Levin M, Apostolovic D, Biedermann T, Commins SP, Iweala OI, Platts-Mills TAE, et al. Galactose α -1,3-galactose phenotypes. *Ann Allergy Asthma Immunol.* (2019) 122:598–602. doi: 10.1016/j.anai.2019.03.021
22. Elferink H, Rossing E, Huang CH, Lee BW, Cao L, Delsing DJ, et al. Modular synthesis and immunological evaluation of suspected allergenic galactooligosaccharides. *Org Biomol Chem.* (2019) 17:3108–12. doi: 10.1039/C9OB00108E
23. Kuehn A, Hilger C, Lehnert-Weber C, Codreanu-Morel F, Morisset M, Metz-Favre C, et al. Identification of enolases and aldolases as important fish allergens in cod, salmon and tuna: component resolved diagnosis using parvalbumin and the new allergens. *Clin Exp Allergy.* (2013) 43:811–22. doi: 10.1111/cea.12117
24. Ferreira F, Hirtenlehner K, Jilek A, Godnik-Cvar J, Breiteneder H, Grimm R, et al. Dissection of immunoglobulin E and T lymphocyte reactivity of isoforms of the major birch pollen allergen Bet v 1: potential use of hypoallergenic isoforms for immunotherapy. *J Exp Med.* (1996) 183:599–609. doi: 10.1084/jem.183.2.599
25. Hilger C, van Hage M, Kuehn A. Diagnosis of allergy to mammals and fish: cross-reactive vs. specific markers. *Curr Allergy Asthma Rep.* (2017) 17:64. doi: 10.1007/s11882-017-0732-z
26. Radauer C, Nandy A, Ferreira F, Goodman RE, Larsen JN, Lidholm J, et al. Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences. *Allergy.* (2014) 69:413–9. doi: 10.1111/all.12348
27. Waite KJ. Blackley and the development of hay fever as a disease of civilization in the nineteenth century. *Med Hist.* (1995) 39:186–96. doi: 10.1017/S0025727300059834
28. Johnson P, Marsh DG. Allergens from common rye grass pollen (*Lolium perenne*). I. Chemical composition and structure. *Immunochemistry.* (1966) 3:91–100. doi: 10.1016/0019-2791(66)90290-4
29. Waldrop SL, Pitcher CJ, Peterson DM, Maino VC, Picker LJ. Determination of antigen-specific memory/effector CD4+ T cell frequencies by flow cytometry: evidence for a novel, antigen-specific homeostatic mechanism in HIV-associated immunodeficiency. *J Clin Invest.* (1997) 99:1739–50. doi: 10.1172/JCI119338
30. Ball T, Fuchs T, Sperr WR, Valent P, Vangelista L, Kraft D, et al. B cell epitopes of the major timothy grass pollen allergen, phl p 1, revealed by gene fragmentation as candidates for immunotherapy. *FASEB J.* (1999) 13:1277–90. doi: 10.1096/fasebj.13.11.1277
31. Etto T, de Boer C, Prickett S, Gardner LM, Voskamp A, Davies JM, et al. Unique and cross-reactive T cell epitope peptides of the major Bahia grass pollen allergen, Pas n 1. *Int Arch Allergy Immunol.* (2012) 159:355–66. doi: 10.1159/000338290
32. Schenk S, Breiteneder H, Susani M, Najafian N, Laffer S, Duchêne M, et al. T-cell epitopes of Phl p 1, major pollen allergen of timothy grass (*Phleum pratense*): evidence for crossreacting and non-crossreacting T-cell epitopes within grass group I allergens. *J Allergy Clin Immunol.* (1995) 96:986–96. doi: 10.1016/S0091-6749(95)70237-7
33. Chabre H, Gouyon B, Huet A, Baron-Bodo V, Boran-Bodo V, Nony E, et al. Molecular variability of group 1 and 5 grass pollen allergens between Pooideae species: implications for immunotherapy. *Clin Exp Allergy.* (2010) 40:505–19. doi: 10.1111/j.1365-2222.2009.03380.x
34. Nony E, Timbrell V, Hrabina B, Boutron M, Solley G, Moingeon P, et al. Specific IgE recognition of pollen allergens from subtropical grasses in patients from the subtropics. *Ann Allergy Asthma Immunol.* (2015) 114:214–20.e2. doi: 10.1016/j.anai.2014.12.005
35. Besche A. On Asthma bronchiale in Man provoked by cat, dog and different other animals. *Acta Med Scand.* (2009) 92:237–55. doi: 10.1111/j.0954-6820.1937.tb13652.x
36. Kogevinas M, Zock JP, Jarvis D, Kromhout H, Lillienberg L, Plana E, et al. Exposure to substances in the workplace and new-onset asthma: an international prospective population-based study (ECRHS-II). *Lancet.* (2007) 370:336–41. doi: 10.1016/S0140-6736(07)61164-7
37. Raulf M. Allergen component analysis as a tool in the diagnosis and management of occupational allergy. *Mol Immunol.* (2018) 100:21–7. doi: 10.1016/j.molimm.2018.03.013
38. Lopata AL, Jeebhay MF. Airborne seafood allergens as a cause of occupational allergy and asthma. *Curr Allergy Asthma Rep.* (2013) 13:288–97. doi: 10.1007/s11882-013-0347-y
39. Meseguer Arce J, Villajos IM, Iraola V, Carnés J, Fernández Caldas E. Occupational allergy to aquarium fish food: red midge larva, freshwater shrimp, and earthworm. A clinical and immunological study. *J Invest Allergol Clin Immunol.* (2013) 23:462–70.
40. Lopata AL, Fenemore B, Jeebhay MF, Gäde G, Potter PC. Occupational allergy in laboratory workers caused by the African migratory grasshopper *Locusta migratoria*. *Allergy.* (2005) 60:200–5. doi: 10.1111/j.1398-9995.2005.00661.x
41. Jeebhay MF, Moscato G, Bang BE, Folletti I, Lipinska-Ojrzanowska A, Lopata AL, et al. Food processing and occupational respiratory allergy-A EAACI position paper. *Allergy.* (2019). doi: 10.1111/all.13807. [Epub ahead of print].

Conflict of Interest: AP is employed by the company Indoor Biotechnologies, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Chan, Pomés, Hilger, Davies, Mueller, Kuehn, Lopata, Gadermaier, van Hage, Raulf and Goodman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.