REVIEW

When does a protein become an allergen? Searching for a dynamic definition based on most advanced technology tools

A. Mari*†

**Center for Clinical and Experimental Allergology, IDI-IRCCS, Rome, Italy and* †*Allergy Data Laboratories s.c., Latina, Italy*

OnlineOpen: This article is available free online at www.blackwell–synergy.com

Clinical and Experimental Allergy

Correspondence:
Dr Adriano Mari, Center for Clinical and Experimental Allergology, IDI-IRCCS, Via dei Monti di Creta, 104 I-00167 Rome, Italy.
E-mail: adriano.mari@allergome.org

Summary

Since the early beginning of allergology as a science considerable efforts have been made by clinicians and researchers to identify and characterize allergic triggers as raw allergenic materials, allergenic sources and tissues, and more recently basic allergenic structures defined as molecules. The last 15–20 years have witnessed many centres focusing on the identification and characterization of allergenic molecules leading to an expanding wealth of knowledge. The need to organize this information leads to the most important question 'when does a protein become an allergen?' In this article, I try to address this question by reviewing a few basic concepts of the immunology of IgE-mediated diseases, reporting on the current diagnostic and epidemiological tools used for allergic disease studies and discussing the usefulness of novel biotechnology tools (i.e. proteomics and molecular biology approaches), information technology tools (i.e. Internet-based resources) and microtechnology tools (i.e. proteomic microarray for IgE testing on molecular allergens). A step-wise staging of the identification and characterization process, including bench, clinical and epidemiological aspects, is proposed, in order to classify allergenic molecules dynamically. This proposal reflects the application and use of all the new tools available from current technologies.

Introduction

More than a century of allergology has been dedicated to the discovery of allergenic sources that cause IgE-mediated diseases. This has involved a step-by-step process moving from identification of raw material causing allergic reactions (i.e. house dust) to organisms and tissues as triggers (i.e. pollen, fruits, mites, fungi). Raw materials, organisms and their tissues have been used and are still in use for diagnostic and therapeutic purposes as allergenic extracts. Although considerable efforts have been made by manufacturers and authorities to control allergenic extract composition, the best definition for an allergenic extract is still 'an unpredictable mixture of allergenic and non-allergenic compounds'. The application to allergen discovery of new biochemical methods during the late 1970s and the 1980s has led to the identification of the real primary sensitizer and trigger, the allergenic molecule. A further spike in allergenic molecule research has been brought about by the progressive and rapid introduction of molecular biology techniques into this research field. Neither the environmental allergenic source identification process nor the characterization of allergenic

molecules has reached a plateau phase, the former being a consequence of both the increasing exposure to novel organisms or the increasing awareness of allergy as the cause of symptoms, and the latter being the consequence of an increasing number of research centres working on allergenic molecule identification and characterization. Both processes are further influenced by an increasing world-wide interest in the field of allergic diseases, mostly in emerging countries. Such a historical trend is readily depicted by monitoring and reporting the number of newly identified allergenic molecules and the number of papers published in the scientific literature from 323 papers in 2000 compared with 870 in 2007 (Fig. 1). This has involved an increasing range of publications from allergy and immunology to biochemical, agricultural and environmental journals. This increase in the knowledge of potentially allergenic molecules requires a systematic organization and a clear definition of the criteria for defining what comprises an allergen, starting from the very first questions: 'what are we going to classify?' or 'which is the structure to be defined as allergenic?' or 'should we consider all the IgE-binding structures?' To address the need to bring sense and organization to the
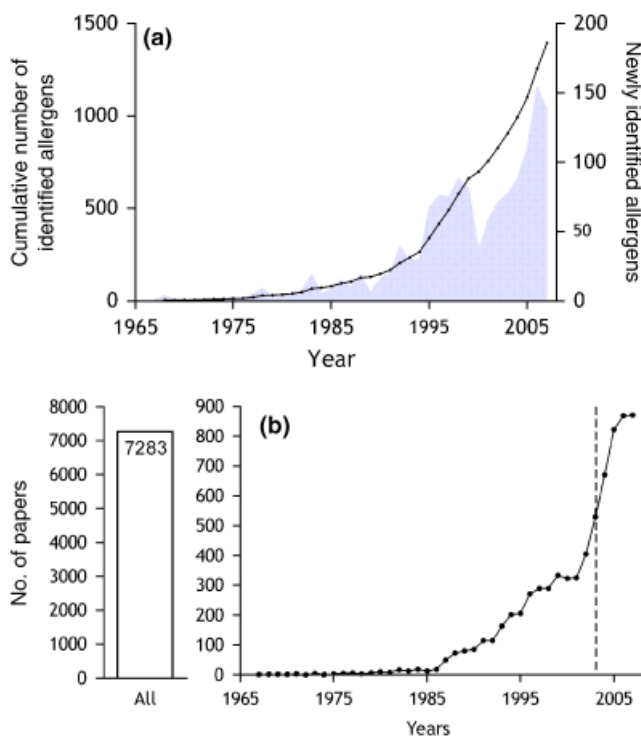
Fig. 1. (a) Last 40 years' time course of new allergen identification reported by either the cumulative number (line) or the newly identified one (shaded area). (b) Last 40 years' time course of published papers reporting on any aspect related to allergenic molecules, as cumulative number (left graph) and by years (right graph). The dashed vertical line indicates Allergome web release (http://www.allergome.org).

increasing amount of data on potential allergens we need to briefly consider some critical aspects of the IgE immune response, report on the current diagnostic and epidemiological tools used for allergic disease studies and the need to implement them and lastly discuss the usefulness of novel biotechnology, information technology and microtechnology tools.

The immunoglobulin E immune response

IgE as a defence provided by the human system seems to be evolutionarily linked to parasite infections [1], not least because of the very high IgE levels generally detected in parasite infection-affected subjects. It is interesting that allergic reactions are rarely described in such patients although they have specific IgE and antigen exposure at the same time. Besides the relevance to infection with helminthic parasites, IgE has been studied because of its role in causing allergic diseases. Such an IgE response differs markedly from the response to parasite antigens as it does not protect from any infective agent and can instead cause disease. Thus, we have the same antibody recognizing different structures depending on the involved organism and playing or not playing a pathogenic role. This distinction is not exclusive to the dichotomy

between parasite and allergenic organisms, but also exists within the IgE recognition of allergenic sources themselves. It is now accepted that IgE from allergic individuals can recognize antigenic non-parasite-related structures from allergens and non-allergens without any related symptoms [2]. A paradigmatic example is IgE recognition of glycan side chains of glycoproteins [3]. Such IgE seems to be unable to trigger allergic symptoms as shown by the negative double-blind placebo-controlled oral challenge with recombinant human lactoferrin produced in rice as a model of plant multiple glycosylation IgE-binding structure [4]. Overall, the condition of having IgE with or without clinical symptoms can be reported for almost any allergen [2].

At the end of this first section, we can assume that IgE can recognize allergenic and non-allergenic structures and may or may not trigger allergic symptoms. Compared with other human diseases, where two conditions are recorded as the absence or the presence of a symptomatic disease (i.e. diabetes, hypertension), measured by the alteration of a single parameter (i.e. blood glucose, blood pressure), if a general population is examined we might record three conditions by using specific IgE detection: (1) no specific IgE; (2) specific IgE sensitization to a given allergen without any clinical symptoms upon exposure; and (3) specific IgE sensitization with different severity degrees of clinical symptoms upon exposure to different levels of allergenic molecules. These three conditions should be kept in mind when defining a molecule as an allergen along with the assumption that the IgE-binding property is not an intrinsic feature of any protein but it is by definition the interaction of two molecules: the antigen and the IgE.

The diagnostic and epidemiological approach

For routine diagnostic purposes or to define the magnitude of allergic sensitization within a given cohort examined for epidemiological purposes, we currently use the skin prick test (which indirectly shows us the presence of specific IgE) and direct IgE detection on serum samples by several *in vitro* methods. In either case, we never make any conclusion about the clinical relevance of the findings based solely on *in vivo* and *in vitro* tests. We need to carefully explore the patient's history searching for evidence of a reliable temporal relationship between allergen exposure and symptom appearance or the individual reactivity by challenging target organs. In the absence of a positive clinical or challenge history, we conclude that the subject is asymptomatic for that given IgE-binding structure. The immunological dichotomy described above is thus defined by the clinical perspective, but it is only from the same testing procedures that we can start the process of classifying compounds as IgE-binding structures. Historically, it is from evidence that exposure causes allergy symptoms that we start the process of identifying

the involved allergenic molecules. Once sensitization is identified, several steps are required before the protein can be fully documented as an allergen. Table 1 reports an ideal process for a comprehensive definition of the characteristics of an allergenic molecule. If we take as for example some allergens that have been identified 20 years ago or more, like Der p 1 or Bet v 1, we see from the literature that they could reach the +4 evaluation stage nowadays, having only occasional studies in the +5 or the +6 evaluation stages. In an almost 10-year range, Pru p 3 has reached 124 publications with just one epidemiological study, leading the allergen to be scored +4. If we apply the 12-step approach to the allergen identification and characterization process, allergens listed in any online publicly accessible repository would range between the − 1 and +2 stages. Two different approaches are currently identifiable within the allergen characterization process: immunochemical identification lacking any biochemical characterization and the direct cloning of the molecule. The former is mostly applied when a polyclonal, IgG or IgE isotype, from either humans or experimental animals, or a monoclonal antibody are available for an already characterized homologous molecule. If the antibody is purified and the allergen specificity is known, the presence of the allergen can be deduced. Antibodies other than IgE give us just the idea that the molecule is in the extract, rather than information about its allergenic nature. IgE inhibition by a previously purified allergen is another immunochemical method to detect and describe an IgE-binding structure within a given allergenic extract. The direct cloning of the allergen is the second strategy adopted when the amount of molecule recoverable from the allergenic extract is too low for any analytical and preparative use or to speed up

allergenic protein characterization. As reported in Table 1 for the natural forms, recombinants should undergo the same stepwise process for characterization plus the validation against the natural form, when available. Because of the assumption reported above, about the presence or the absence of clinical disease in allergen-specific IgE-bearing individuals, the definition of the symptomatic/asymptomatic ratio would most likely be defined at stages +5 and +6 of the allergen characterization process.

At the end of this second section, we can conclude that, having an extended although not yet exhaustive knowledge on allergenic molecules, the identification process can also be somewhat different from the 'classical' biochemical/immunochemical one. As reported above and from a survey of the literature, this process is always dispersed on several papers published in more than one journal depending on the leading topic. We can still certainly distinguish an identification process for a very new compound, which is almost unchanged since the very early days of allergen identification and today called 'proteomics', but at the same time several different approaches can be applied to homologous allergen identification by using molecular cloning technique [5]. The stepwise process seems to be a reliable way to collect accumulating data on each allergenic molecule.

Biotechnology, information technology and microtechnology tools

From the first two sections, we have learned that we need to organize our knowledge on structures and let them be defined as IgE binding or not, and as causing allergic diseases or not. Biotechnology procedures, which include

Table 1. Allergen identification and full characterization process

− 5	Suspicious of an allergic reaction to an organism or its tissues
− 4	Preparation of the best extract starting from the best raw material
− 3	Positive skin testing and IgE testing with the extract
− 2	Extract evaluation by SDS-PAGE
− 1	IgE immunoblot identification of SDS-PAGE isolated bands
0	Isolation and preliminary sequencing of the IgE-reactive band(s)
+1	Purification of the identified IgE-reactive band and full biochemical characterization, including source tissue localization and concentration under several physiological and pathological conditions, and molecular cloning of the allergen
+2	Evaluation of the naturally purified molecule by skin testing and IgE binding including any basophil/mast cell activation test (5–15 subjects) (should also apply to the recombinant form)
+3	Evaluation of the naturally purified molecule by *in vivo* challenging of affected and non-affected organs (should also apply to the recombinant form)
+4	Evaluation of the naturally purified molecule on a broader population affected by the same sensitization (extract-detected) (should also apply to the recombinant form)
+5	Evaluation of the IgE reactivity of naturally purified molecule within a general allergic population, in one or more geographical areas (should also apply to the recombinant form). Defining the symptomatic/asymptomatic ratio
+6	Evaluation of the IgE reactivity of naturally purified molecule within a general population, in one or more geographical areas (should also apply to the recombinant form). Defining the symptomatic/asymptomatic ratio

Characterization stages defined by the *minus* sign refer to pre-allergenic structure definition; those defined by the *plus* sign refer to the post-allergenic structure definition.

all the molecular biology techniques, might lead us to potentially identify all the IgE-binding structures regardless of whether they are either rarely expressed in the allergenic organism tissue or are rarely IgE recognized within a general 'average' population, or are rarely but constantly IgE recognized by selected subsets of genetically prone subjects [6]. The advantage of any rapid high-throughput system for allergen identification [7] or even the genomic identification of related genes [8, 9] is to give us access to hundreds of potentially related or putative allergenic molecules. In the case of the 12-step approach described in Table 1, the allergen identification and characterization starts at stage +1, but lacks the definition of the IgE-binding property. We expect to have more of these studies leading to an extended genomic knowledge of known allergenic organisms and taxonomic related and unrelated ones. We must thus implement strategic tools to classify the structures since the beginning of their description by extensively using information technology tools. At the same time, we must step up the characterization stage of every single molecule using high-throughput IgE detection systems. There are several Internet-based resources aiming to regulate and classify IgE-binding structures [10–12]. Three will be analysed herein as they represent a different approach to allergen classification. The Allergen Nomenclature web site (http://www.allergen.org) is devoted to report the work of the specific WHO-IUIS subcommittee. Being an official organ of an international health authority organism, the expert work of the subcommittee should be advised as the most critical one in a blasting situation of allergen identification as reported in Fig. 1. Following predefined rules [13], names are given to submitted protein identified as allergens. Fulfilling the given criteria, the new allergen receives a name and is put in the list. No additional data are collected thereafter. Looking at the last WHO-IUIS-released allergen list (28 January 2008), 580 allergens and 876 isoallergens are classified. It is noteworthy that if not submitted, many well-characterized allergens are not in the WHO-IUIS list (i.e. buckwheat allergens). The allergen recruitment mechanism is the basis of the discrepancy between the WHO-IUIS allergen list and the others. The overall data processing sometimes leads to allergens classified before any public document appears in the scientific literature. Only one reference paper is given for each allergen. No computational tools are available from the web site.

The Food Allergy Research & Resource Program (FAARP) at the University of Nebraska releases a web-based database (www.allergenonline.com) whose aim is to classify proteins as allergens. This work is based on a panel of scientists and clinicians actively involved in reviewing literature data by comparing peer-reviewed publications following predetermined guidelines. Pre-defined criteria lead to classify proteins into three categories as allergens, putative allergens and those having insuffi-cient evidence of allergenicity. The last released database, version 8.0, lists 1313 peer-reviewed sequences included in the first two categories. Allergenicity or putative allergenicity of a given sequence is evaluated, but assignment to the two categories is not reported in the accessible database. Further, considering the given criteria for allergenicity (i.e. 'specifically bind IgE using sera from individuals with clear allergies to the source of the gene/protein and further that the protein causes basophil activation or histamine release, skin test reactivity or challenge test reactivity using subjects allergic to the source'), many of the listed isoforms should not be considered either allergens or putative allergens as for instance they have never been shown to bind IgE. Reference documentation is given for allergen groups generating confusion when reaching a putative allergen sometimes having just a sequence similarity to a given allergen, thus seeming to be well documented. The availability of sequence similarity search tools, either customized or not, increases the usefulness of the database, particularly if the main purpose of the database is to evaluate allergenicity of novel proteins from genetically modified organisms.

The Allergome platform (http://www.allergome.org) is an independent project aiming to classify allergens, IgE-binding antigens and non-IgE-binding structures. The only pre-defined criterion used for the entry of a new structure is that it has to be tested at least once for its IgE-binding capacity or has any structural relationship with known allergens. This strategy leads to classify the largest set of molecules among online available web sites (1400 allergens, 939 isoallergens, as of February 2008). Each classified structure is fully documented in terms of published papers, each study is assigned to a specific scientific documentation category (i.e. biochemistry, molecular biology, diagnosis, epidemiology, etc.), allowing the user to understand whether the molecule has just a sequence reported in protein databases or has been fully documented, somewhat helping to comply with the classification stage as described in Table 1. References may be retrieved either by searching for the specific molecule or by the Allergome customized literature mining tool, the RefArray. Data on allergenicity are also extracted from the literature and reported in detail in the molecule monograph. At the moment, no search tools are available to discriminate between molecules on the basis of the documentation extent. A tool allowing dynamic refined searches among classified structures would certainly increase the usefulness of the Allergome platform, fulfilling the need and requirements for a more critical view of allergenic structure classification and characterization level. No computational tools are available from the Allergome web site. None of the Internet-based databases has adopted any post-identification criteria for a further characterization of a given allergenic molecule.

Allergens should undergo an extensive evaluation based on their use during the clinical workup in world-wide allergy centres or within large epidemiological studies. Dealing with hundreds of allergenic molecules would be difficult or almost impossible using skin test or singleplex testing currently available in laboratory systems to support this idea. Following the application of microarray technology to genomics, the same technology is now applied to proteomics. A large collaborative study involving world-wide researchers gave us the proof of concept that proteomic microarray can be dedicated to IgE detection based on allergenic molecules [14]. Further reports describe this new powerful tool [15–17] now available as a commercial product for routine multiplex IgE detection (ISAC system, VBC-Genomics, Vienna, Austria). The intrinsic feature of the allergenic molecule-based multiplex test is to allow any single allergy centre adopting the tool for routine diagnostic purposes to produce a wealth of data related to all the tested molecules on all the tested patients. This molecule-based microarray IgE testing, adopted in our centre since March 2006, has generated almost 2 000 000 data on an average of 80 allergens, giving a detailed magnitude of sensitization prevalence for each allergen and linking them to clinical features. Further expansion of available allergenic molecules spotted on the microarray, regardless of being common or uncommon allergens in a given geographical area, would further expand our knowledge. It is noteworthy that additional powerfulness will be added to such testing when the ISAC system will be fully interfaced with the Allergome-ReTiME platform, a module for raw data mining allowing to store and retrieve real-time produced data. Complementing these networked systems with an electronic record for the allergy patient would allow any world-wide allergy specialist to contribute towards increasing the knowledge on allergens. Additional bioinformatics analysing tools will help in evaluating protein reciprocal biochemical and immunochemical relationships by showing for instance the clustering behaviour of previously undefined structure.

Conclusions

It is highly realistic to say that the definition of the biological feature of any given protein to be or not an IgE-binding protein, and to be capable of giving rise to clinical symptoms in a certain percentage of the sensitized population, comes from a stepwise approach combining bench and bedside studies, starting from still relevant isolated case reports to selected cohort studies. Such 'classical' tools can now be integrated by biotechnology, microtechnology and bioinformatics tools in an attempt to dynamically describe the allergen/non-allergen ontology. The stepwise approach for defining a molecule as an allergen or not may require months to years, but a dynamically defined platform integrating the best feature of currently available web-based resources could speed up our process to acquire the most comprehensive knowledge on allergenic structures ever, leading most of the allergenic molecules to step up their characterization level in a short time frame, and to be classified between the highest stages of the arbitrary scale reported in Table 1. In the future, considerable effort should be directed towards designing the broadest world-wide project in order to classify IgE- and non-IgE-binding structures using most of the technological resources as reported in the present article and involving health authorities, scientific associations and companies. Although it could seem too visionary a hypothesis for the future, too far from a realistic application, the opinions reported herein are based on their development and their practical current application.

currently the main independent project financially supported by Allergy Data Laboratories s.c.

References

1 Maizels RM, Yazdanbakhsh M. Immune regulation by helminth parasites: cellular and molecular mechanisms. *Nat Rev Immunol* 2003; **3**:733–44.

2 Bousquet J, Anto JM, Bachert C *et al.* Factors responsible for differences between asymptomatic subjects and patients presenting an IgE sensitization to allergens. A GA²LEN project. *Allergy* 2006; **61**:671–80.

3 van Ree R. Carbohydrate epitopes and their relevance for the diagnosis and treatment of allergic diseases. *Int Arch Allergy Immunol* 2002; **129**:189–97.

4 Mari A, Ooievaar-de Heer P, Scala E *et al.* Evaluation by double-blind placebo-controlled oral challenge of the clinical relevance of IgE antibodies against plant glycans. *Allergy* 2008; DOI:10.1111/j.1398-9995.2008.01703.x.

5 Rhyner C, Weichel M, Fluckiger S, Hemmann S, Kleber-Janke T, Crameri R. Cloning allergens via phage display. *Methods* 2004; **32**:212–8.

6 Nikolaizik WH, Weichel M, Blaser K, Crameri R. Intracutaneous tests with recombinant allergens in cystic fibrosis patients with allergic bronchopulmonary aspergillosis and Aspergillus allergy. *Am J Respir Crit Care Med* 2002; **165**:916–21.

7 Kodzius R, Rhyner C, Konthur Z *et al.* Rapid identification of allergen-encoding cDNA clones by phage display and high-density arrays. *Comb Chem High Throughput Screen* 2003; **6**:147–54.

8 Hong SG, Cramer RA, Lawrence CB, Pryor BM. Alt a 1 allergen homologs from Alternaria and related taxa: analysis of phylogenetic content and secondary structure. *Fungal Genet Biol* 2005; **42**:119–29.

9 Gao ZS, van de Weg WE, Schaart JG *et al.* Genomic cloning and linkage mapping of the Mal d 1 (PR-10) gene family in apple (*Malus domestica*). *Theor Appl Genet* 2005; **111**:171–83.

10 Schein CH, Ivanciuc O, Braun W. Bioinformatics approaches to classifying allergens and predicting cross-reactivity. *Immunol Allergy Clin North Am* 2007; **27**:1–27.

11 Gendel SM, Jenkins JA. Allergen sequence databases. *Mol Nutr Food Res* 2006; **50**:633–7.

12 Mari A. Importance of databases in experimental and clinical allergology. *Int Arch Allergy Immunol* 2005; **138**:88–96.

13 WHO/IUIS. Allergen nomenclature. WHO/IUIS Allergen Nomenclature Subcommittee World Health Organization, Geneva, Switzerland. *Clin Exp Allergy* 1995; **25**:27–37.

14 Hiller R, Laffer S, Harwanegg C *et al.* Microarrayed allergen molecules: diagnostic gatekeepers for allergy treatment. *FASEB J* 2002; **16**:414–6.

15 Harwanegg C, Hiller R. Protein microarrays for the diagnosis of allergic diseases: state-of-the-art and future development. *Clin Chem Lab Med* 2005; **43**:1321–6.

16 Harwanegg C, Hiller R. Proteinmicroarrays in diagnosing IgE-mediated diseases: spotting allergy at the molecular level. *Expert Rev Mol Diagn* 2004; **4**:539–48.

17 Harwanegg C, Laffer S, Hiller R *et al.* Microarrayed recombinant allergens for diagnosis of allergy. *Clin Exp Allergy* 2003; **33**:7–13.