HISTORY OF CHARACTERIZING ALLERGENS

Ronald van Ree
Academic Medical Center
University of Amsterdam
Von Pirquet chose this word to illustrate that the immune system does not react with “normal immunity” to a foreign protein but with hyper-sensitivity.
What is allergy?

In principle there are 4 types:

1. **Type I: IgE antibody-mediated**
   - Rhino-conjunctivitis (hay fever)
   - Allergic asthma
   - Allergic eczema/dermatitis
   - Food allergy
Subsequent steps in the development of type I allergy

1. Antigen presentation to naive T-cells:
   - innate signals
   - dose
   - genetic predisposition
   → allergen-specific Th2 cells

2. Contact allergen-specific B-cell and Th2: class-switch to IgE production
   → allergen-specific IgE

3. Sensitisation of effector cells: binding of IgE to high-affinity IgE receptor

4. Repeated contact with allergen: cross-linking of IgE resulting in mediator release by effector cells, e.g. mast cells

5. Immediate and late-phase reactions
IgE-mediated allergic reaction
Mediator release by mast cells
What is an allergen molecule?

1. An allergen is a molecule that binds IgE antibodies
   If a molecule can not bind IgE it is no allergen

2. An allergen is a molecule that induces IgE antibodies
   (and consequently also binds IgE antibodies)
   This is certainly not the case for all molecules that bind IgE

3. An allergen is a molecule that induces allergic symptoms
   Again, this is not true for all molecules that bind IgE

Food allergy is the best example to illustrate the differences
CROSS-REACTIVITY

Cross-reactivity is a property attributed to IgE antibodies when they recognize homologous molecules from other sources than the one that originally induced IgE.
Structural homology is the explanation of cross-reactivity

“Homology to birch pollen Bet v 1 makes the antigen Pru av 1 an allergen but NOT an allergen that was responsible for the induction of IgE”
Most stringent definition:

“An antigen that sensitizes (induces IgE) and (usually) causes symptoms”

COMPLETE ALLERGEN

Cross-reactive allergen:

“An antigen that does/can not sensitize itself but can cause symptoms”

INCOMPLETE ALLERGEN

Ergo, there are two potential risks when introducing a transgene:

• Introducing a risk for de novo sensitization (induction of IgE)

• Introducing a risk for inducing symptoms via pre-existing (cross-reactive) IgE
What is the origin of food sensitization and allergy?

There are essentially two ways to become sensitized and/or food allergic:

1. Exposure to foods such as egg, milk, fish, peanut or hazelnut

2. Exposure to respiratory allergens such as pollen or mites

   cross-reactivity to foods
Some examples

COMPLETE ALLERGENS (sensitizers/ frequent symptoms):

**Inhalant:** Bet v 1, Phl p 1, Phl p 5, Der p 1, Der p 2, Fel d 1, etc.

**Food:** Ara h 2, Pru p 3, Cor a 9, Gad c 1, Gad m 1, etc.

INCOMPLETE ALLERGENS (cross-reactive / possible symptoms):

**Food:** Bet v 1-homologues in fruits, nuts and vegetables

profilin homologues in virtually all plant foods

plant N-glycans

Many cross-reactivities are not clinically relevant, at least not in every patient. Sometimes no evidence for clinical relevance can be found at all.
Typical IgE-binding plant N-glycans

- GlcNAc - GlcNAc - Man
  |    |    |
  Fuc  |  Xyl  | Man
  |     | Man   |
Expression of human lactoferrin in rice: plant glycosylation

primate glycosylation

plant glycosylation
Expression of human lactoferrin in rice introduces IgE-binding glycans.
DBPCOC with rice lactoferrin in 5 plant glycan-IgE positive subjects

All five subjects had a negative challenge with a cumulative dose of 1.111 gram of transgenic lactoferrin

*Mari A. et al. Allergy 2008; 63: 891-6*
Lesson from the plant glycan specific IgE antibodies:

Not all IgE binding has clinical relevance

Not all IgE binding increases the risk of allergic reactions
Back to history: early description of food allergy

QUOTE: “For cheese does not prove equally injurious to all men, for there are some who can take it to satiety, without being hurt by it in the least, but, on the contrary, it is wonderful what strength it imparts to those who agree it agrees with; but there are some who do not eat it well, their constitutions are different, they differ in this respect, that what in their body is incompatible with cheese, is roused and put in commotion by such a thing; and those in whose bodies such a humor happens to prevail in greater quantity and intensity, are likely to suffer the more from it. But if the thing had been pernicious to the whole nature of man, it would have hurt all.”

460 BC – 370 BC
What about respiratory allergy?

Why is this girl sneezing?
John Bostock (London 1773 – 1846)

First physician to accurately describe hay fever in 1819.

John Elliotson (London 1791 – 1868)

First physician to link hay fever to pollen of grass in 1831
It took more than 100 years longer to discover the cause of most allergic asthma cases worldwide.
Another major breakthrough for the field: discovery of IgE

Kimishige and Teruko Ishizaka

1967

Gunnar Johansson and Hans Bennich
First allergen molecules to be identified and immunochemically characterized were from ragweed, perennial rye grass and cod fish.

This was around the end of the sixties and beginning of the seventies of the 20\textsuperscript{th} century.

One more thing was needed: molecular cloning. The biotechnology revolution started in the 70s.
First allergen was cloned and expressed in 1988 by Wayne Thomas from Perth/Australia:

Der p 1 from *Dermatophagoides pteronyssinus*

This was followed shortly by Dol m 5 (wasp) cloned by T.P. King.

In 1989, Der p 2, Der p 5, Bet v 1 and Dac g 2 followed.

Bet v 1 was cloned by Heimo Breiteneder and was the first recombinant allergen for which IgE binding was demonstrated.
From the 90s up to the present day hundreds and hundreds of allergens have been identified, cloned, sequenced, expressed and characterized.

The most important allergens (major allergens) of the most important allergen sources have by now been identified and obtained as recombinant reagents.

This has facilitated to evaluate whether proteins have particular properties that make them into allergens.
Some examples of allergens where their functional or physico-chemical properties have been proposed to favour their role as allergens.
Allergenicity resulting from functional mimicry of a Toll-like receptor complex protein

Aurelien Trompette et al.

Nature 2008 Dec 7
Pro-inflammatory cytokines (IL6, TNF,...)

Events leading to triggering the TLR4 signaling cascade

Lipoplysaccharide-binding protein (LBP) acts as a lipotransferase (produced in liver/lungs)

Binding of LPS to MD2/TLR4 complex results in dimerization of two of these complexes (heterotetramer) and triggers off the signaling cascade.

Similarity in folding: two anti-parallel \(\beta\)-pleated sheets (immunoglobulin superfamily) surrounding a hydrophobic cleft


From: Derewenda U et al. The crystal structure of a major house dust mite allergen Der p 2, and its biological implications. J. Mol. Biol. 2002; 318:189-197
Why is Der p 2 a major allergen?

Or perhaps more generally, why are major allergens major allergens? What make a protein a strong allergen?

The paper supports that the origin of Der p 2’s strong allergenicity resides in it’s mimicry of MD2, thereby functionally synergizing with (or even replacing) MD2 in LPS-driven triggering of TLR4 signaling driving Th2 immunity.

Or even more generally, that this is a general property of many major allergens because at least 50% has lipid-binding properties.
Two other lipid-binding allergens reveal another property that has impact on allergenicity: protease resistance.

Bet v 1 from birch pollen: cross-reactivity to fruits

LTP (lipid transfer protein) from fruits: complete allergen
A European study on apple allergy

IgE testing using four purified apple allergens:

1. Birch-pollen cross-reactive allergen
2. Mal d 2
3. Lipid transfer protein (LTP): true food allergen
4. Mal d 4

There is a clear geographic difference. So what?
Only patients with IgE antibodies against the COMPLETE ALLERGEN have severe systemic symptoms (U: generalized urticaria / AX: anaphylaxis). The risk is for severe food allergy increased by around 8-fold!

Likely explanation: resistance to gastric digestion

IgE to a transgene with homology to a protease-resistant allergen may bear more risk than to a protease-sensitive allergen.
Allergenicity grape LTP unaffected by gastric/duodenal digestion

confirmed by SPT

Vassilopoulou et al. JACI 2006;118:473-480
Several major allergens have been classified as proteases:

• Group 1 house dust mite allergens are cysteine proteases (e.g. Der p 1 and Der f 1)
• Group 3, 6, 9 house dust mite allergen are serine proteases
• Cockroach aspartic proteinases (e.g. Bla g 2)
• Cockroach group 1 allergens (e.g. Bal g 1)?
• Pollen proteases?
• Mould serine proteases

So what are the postulated mechanisms by which proteases facilitate induction of sensitization and allergic symptoms?
Most data in support of proteolytic action as an answer to the question “what makes an allergen an allergen?” come from studies on the group 1 house dust mite allergens Der p and f 1.
Three main targets of Der p 1/Der f 1’s protease activity:

• physical barrier function

• innate responses of structural cells (epithelial cells / keratinocytes)

• adaptive immune response
Physical barrier function

Der p 1 facilitates transepithelial allergen delivery by disruption of tight junctions.
Wan H, Winton HL, Soeller C, Tovey ER, Gruenert DC, Thompson PJ, Stewart GA, Taylor GW, Garrod DR, Cannell MB, Robinson C.
J Clin Invest 1999; 104:123-133
There is certainly quite some evidence from *ex vivo, in vitro* and mouse model experiments that there is a potential pro-allergenic pathway triggered off by proteolytic action. Whether this is also the case in vivo in human subjects is hard to say.
What makes a protein a COMPLETE ALLERGEN?

Complex issue:
• *intrinsic factors*, i.e. molecular properties such as:
  - protease activity
  - lipid-binding
  - protease resistance
  - heat resistance
  - affinity for PRRs (e.g. TLRs)

• *extrinsic factors*, i.e. anything “accompanying” the molecule such as:
  - matrix presenting the molecule
  - co-exposures in environment
  - infections
  - microbiome
  - level of exposure
  - genetic background
  - timing and frequency of exposure
Abundance

Storage proteins like 2S, 7S and 11S are all dominant proteins and important allergens.
Abundance of an inhalant look-alike, i.e. cross-reactivity

Bet v 1 is the most dominant protein in birch pollen extract

Bet v 1 causes sensitization by the respiratory route. Profilin as well but only in high pollen exposure areas in highly sensitive individuals.

Abundance is not a requirement for the cross-reactive foods like apple.
Minor allergens of olive mainly recognized in high-pollen exposure area

Overall, for food allergens there is no evidence that specific structural features or biological functions determine allergenicity, with the exception of stability.

Allergenicity of food proteins seems to be the sum of abundance and stability.

Stability is certainly influenced by the food matrix.

Of course it can not be excluded that other factors like enzymatic activity play a role, but so far there is no clear evidence for this.
It is important to realize that we do not have any validated method to predict allergenicity in the sense of the sensitizing potential of a candidate transgene.
Editors’ choice articles

Allergens are distributed into few protein families and possess a restricted number of biochemical functions

Christian Radauer, PhD,a Merima Bublin, PhD,a Stefan Wagner, PhD,a Adriano Mari, MD,b,c and Heimo Breiteneder, PhDa Vienna, Austria, and Latina and Rome, Italy

Conclusion: The small number of protein families that contain allergens and the narrow functional distribution of most allergens confirm the existence of yet unknown factors that render proteins allergenic. (J Allergy Clin Immunol 2008;121:847-52.)

Clinical implications: The classification of allergens supports the elucidation of factors that make proteins allergenic, thus possibly paving the way for novel therapeutic concepts.
Characteristic motifs for families of allergenic proteins

Ovidiu Ivanciuca,b, Tzintzuni Garciam, Miguel Torresec, Catherine H. Scheina,b,d,e, Werner Brauna,b,e,*

a Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555–0857, United States
b Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555–0857, United States
c McAllen High School, McAllen, TX, United States
d Department of Microbiology and Immunology, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555–0857, United States
e Sealy Center for Vaccine Development, University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77555–0857, United States

Allergens group to < 1.5% of > 9000 protein families listed in the Pfam database
## Most abundant Pfam allergen families

<table>
<thead>
<tr>
<th>Pfam code</th>
<th>Pfam domain</th>
<th>No allergens</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF00234</td>
<td>Protease inhibitor/seed storage/LTP family</td>
<td>34</td>
</tr>
<tr>
<td>PF00235</td>
<td>Profilin</td>
<td>27</td>
</tr>
<tr>
<td>PF00036</td>
<td>EF hand</td>
<td>23</td>
</tr>
<tr>
<td>PF01357</td>
<td>Pollen allergen</td>
<td>19</td>
</tr>
<tr>
<td>PF00188</td>
<td>SCP-like extracellular protein</td>
<td>19</td>
</tr>
<tr>
<td>PF00407</td>
<td>Pathogenesis-related protein Bet v 1 family</td>
<td>16</td>
</tr>
<tr>
<td>PF00261</td>
<td>Tropomyosin</td>
<td>16</td>
</tr>
<tr>
<td>PF00190</td>
<td>Cupin</td>
<td>15</td>
</tr>
<tr>
<td>PF00061</td>
<td>Lipocalin/cytosolic fatty-acid binding protein family</td>
<td>12</td>
</tr>
<tr>
<td>PF03330</td>
<td>Rare lipoprotein A (RlpA)-like double-psi beta-barrel</td>
<td>12</td>
</tr>
<tr>
<td>PF00042</td>
<td>Globin</td>
<td>9</td>
</tr>
<tr>
<td>PF00544</td>
<td>Pectate lyase</td>
<td>9</td>
</tr>
<tr>
<td>PF00112</td>
<td>Papain family cysteine protease</td>
<td>8</td>
</tr>
<tr>
<td>PF00428</td>
<td>60s Acidic ribosomal protein</td>
<td>8</td>
</tr>
<tr>
<td>PF00082</td>
<td>Subtilase family</td>
<td>7</td>
</tr>
<tr>
<td>PF00314</td>
<td>Thaumatin family</td>
<td>7</td>
</tr>
<tr>
<td>PF01190</td>
<td>Pollen proteins Ole e 1 family</td>
<td>7</td>
</tr>
<tr>
<td>PF01620</td>
<td>Ribonuclease (pollen allergen)</td>
<td>7</td>
</tr>
<tr>
<td>PF00012</td>
<td>Hsp70 protein</td>
<td>6</td>
</tr>
<tr>
<td>PF00578</td>
<td>AhpC/TSA family</td>
<td>6</td>
</tr>
<tr>
<td>PF00089</td>
<td>Trypsin</td>
<td>5</td>
</tr>
<tr>
<td>PF00113</td>
<td>Enolase, C-terminal TIM barrel domain</td>
<td>5</td>
</tr>
<tr>
<td>PF00187</td>
<td>Chitin recognition protein</td>
<td>5</td>
</tr>
<tr>
<td>PF03952</td>
<td>Enolase, N-terminal domain</td>
<td>5</td>
</tr>
<tr>
<td>PF00151</td>
<td>Lipase</td>
<td>4</td>
</tr>
<tr>
<td>PF00197</td>
<td>Trypsin and protease inhibitor</td>
<td>4</td>
</tr>
<tr>
<td>PF00295</td>
<td>Glycosyl hydrolases family 28</td>
<td>4</td>
</tr>
<tr>
<td>PF01630</td>
<td>Hyaluronidase</td>
<td>4</td>
</tr>
<tr>
<td>Platform</td>
<td>Details</td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>IUIS</td>
<td>lists official names</td>
<td></td>
</tr>
<tr>
<td></td>
<td>links to GenBank, UniProt, PDB</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.allergen.org">http://www.allergen.org</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDAP</td>
<td>allergen sequences, structures, IgE epitopes</td>
<td></td>
</tr>
<tr>
<td>Structural Database of Allergenic Proteins (SDAP)</td>
<td>links to GenBank, UniProt, PDB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FASTA search in SDAP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LAST search in GenBank, SwisProt, PIR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pfam classification</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAO/WHO allergenicity rules</td>
<td></td>
</tr>
<tr>
<td></td>
<td>tools for sequence and epitope comparison</td>
<td></td>
</tr>
<tr>
<td></td>
<td>high-quality allergen models</td>
<td></td>
</tr>
<tr>
<td><a href="http://fermi.utmb.edu/SDAP">http://fermi.utmb.edu/SDAP</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FARRP</td>
<td>allergen list</td>
<td></td>
</tr>
<tr>
<td>Food Allergy Research and Resource Program (FARRP)</td>
<td>FASTA search in FARRP</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAO/WHO allergenicity rules</td>
<td></td>
</tr>
<tr>
<td></td>
<td>links to GenBank</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.allergenonline.org">http://www.allergenonline.org</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allergome</td>
<td>allergen list</td>
<td></td>
</tr>
<tr>
<td></td>
<td>links to PubMed, Uniprot, PDB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>links to sequence databases</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.allergome.org">http://www.allergome.org</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AllFam</td>
<td>Pfam classification</td>
<td></td>
</tr>
<tr>
<td><a href="http://www.meduniwien.ac.at/allergens/allfam">http://www.meduniwien.ac.at/allergens/allfam</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSL (Central Science Laboratory, UK)</td>
<td>allergen list</td>
<td></td>
</tr>
<tr>
<td></td>
<td>links to GenBank</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IgE epitopes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>clinical data</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://allergen.csl.gov.uk/index.htm">http://allergen.csl.gov.uk/index.htm</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>InformAll</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://foodallergens.ifr.ac.uk">http://foodallergens.ifr.ac.uk</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADFS (Allergen Database for Food Safety)</td>
<td>allergen list</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sequences</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IgE epitopes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3D structures</td>
<td></td>
</tr>
<tr>
<td></td>
<td>epitope search</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAO/WHO allergenicity rules</td>
<td></td>
</tr>
<tr>
<td></td>
<td>motif-based allergenicity prediction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>allergen list species description</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://allergen.nihs.go.jp/ADFS/">http://allergen.nihs.go.jp/ADFS/</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Allergy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://allallergy.net">http://allallergy.net</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IEDB</td>
<td>T-cell and B-cell epitopes</td>
<td></td>
</tr>
<tr>
<td>Immune Epitope Database and Analysis Resource (IEDB)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://immuneepitope.org">http://immuneepitope.org</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allermatch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://www.allermatch.org">http://www.allermatch.org</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What can we do with (some) allergen databases?

Evaluate potential risk for inducing symptoms in already sensitized subjects based on degree of sequence homology with known allergens.

Demonstration of specific IgE is in general an alarm bell. The example of plant N-glycans however demonstrates that this may sometimes be overcautious and result in discarding perfectly safe proteins with N-glycosylation sites.

Having said that searching for homology (BIOINFORMATICS) with existing allergens is of great importance to rule out that existing allergens are introduced, or proteins with likely cross-reactivity to existing proteins.
What can we NOT do with (any) allergen database?

Evaluate the risk of de-novo sensitization based on allergen sequences.

Intrinsic pro-allergenic properties can not be more than one of a multitude of factors that in the end determine whether somebody develops an allergy.

Not everybody exposed to a protein with pro-allergenic properties will develop IgE antibodies against it, let alone develop a clinical allergy.

Even if family background, westernized sedentary lifestyle, and early exposure to allergen with reported pro-allergenic properties (e.g. Der p 2) come together, the exposed subject can stay healthy or become allergic to pollen instead of mite.
University of Nebraska Allergen Database
(Allergen Online)

- Industry sponsored, peer-reviewed allergen database at University of Nebraska
  - Peer-reviewed by clinical and research allergists from around the world: Japan, Europe, and USA
  - Well-defined criteria; posted on database website.
  - Inclusion of protein allergens (food, dermal, respiratory) based on available data in the public literature.
  - Updated once a year
  - Available free to the general public

www.allergenonline.org
Allergen Search Strategy

- Compare amino acid sequence of query protein to database containing sequences of food, dermal and respiratory allergens.

• Evaluate sequence for amino acid identity using local alignment programs, such as BLAST (or FASTA)
  • > 35% identity over an 80 or greater amino acid window

and potential (theoretical) IgE epitope matches.
Specific IgE Sera Screening

• For proteins originating from an allergenic source, or having significant homology with a known allergen, specific serum screening is conducted.

• An issue of critical importance to sera screening is the availability of *well characterized*, quality human sera from a sufficient number of patients.

• Potential false positives/equivocal results (e.g. N-glycans)

• Risk assessment: risk for mild versus severe symptoms?
Concluding remarks

• Over the past 4 decades most important allergens have been identified, sequenced and characterized.
• Peer-reviewed databases that are regularly updated are freely accessible to allow sequence comparison of candidate transgenes
• Bio-informatics thus allows preventing introduction of known allergens or cross-reactive molecules that would potentially cause allergic reactions in sensitized individuals
• Currently, no validated methodology is available to predict whether a protein will cause de-novo sensitization. Only protease resistance and abundance of a candidate transgene are potentially relevant parameters to monitor.