Allergen Databases

Steven M. Gendel, Ph.D.
Food and Drug Administration
Center for Food Safety and
Applied Nutrition

Bioinformatics of Allergens

- **The Present**
  - Allergenicity Assessment
    - Methods have become more sophisticated
    - Information on protein structure is being integrated
- **The Future**
  - Understanding the effects of processing
  - Informing decisions related to regulatory thresholds
  - Understanding sensitization

Allergen Databases

- ## Historically
  - Each database developed to support a particular analysis

- ## Therefore
  - Contents vary between databases
  - Accessibility of content varies
  - Level of descriptive annotation varies

Allergen Databases

- ## Two types
  - **Biologic – include clinical, physiological, and other information**
  - **Molecular – focused on sequences and structures of allergenic proteins**

Allergen Databases - Biologic

- **AllAllergy**
  - http://allallergy.net/
- **Allergome**
  - http://www.allergome.org/
- **Informall**
  - http://www.foodallergens.info/

Allergen Databases - Molecular

◆ **Allergen Database for Food Safety [ADFS]**
  - http://allergen.nihs.go.jp/ADFS/

◆ **Allergen Online [AO]**
  - http://allergenonline.com

◆ **AllerDB [ADB]**
  - http://sdmc.i2r.a-star.edu.sg/Templar/DB/Allergen/

◆ **AllerMatch [AM]**
  - http://www.allermatch.org/

◆ **Bioinformatics for Food Safety [BIFS]**
  - http://www.iit.edu/~sgendel/fa.htm

◆ **Central Science Laboratory [CSL]**
  - http://www.csl.gov.uk/allergen/

◆ **Structural Database of Allergen Proteins [SDAP]**
  - http://fermi.utmb.edu/SDAP/sdap_ver.html

Allergen Databases - Other

- **Allergenicity prediction based on calculated motifs/peptides**
  - Evaller
    http://bioinformatics.bmc.uu.se/evaller.html
  - WebAllergen
    http://weballergen.bii.a-star.edu.sg/
  - AllerTool
    http://research.i2r.a-star.edu.sg/AllerTool/
  - AlgPred
    http://www.imtech.res.in/raghava/algpred/

Database Comparisons

Database	Last Update	Criteria Specified?	No. of Sequences
ADFS	Jul 2006	Y	1366
AO	Jan 2007	Y	1251
ADB	?	N	?
AM	Jan 06	Y	792
BIFS	N/A	?	?
CSL	?	N	?
SDAP	Jul 07	Y	829

Database Comparisons

Database	Allergen Categories
ADFS	8
AO	13
ADB	?
AM	1
BIFS	3
CSL	4
SDAP	9

Some User Concerns

- Identification and evaluation of changes in a database
- Development of "third party" applications for comparison or integration
- Need to devote time and resources to developing redundant databases for each new analysis

The Solution?

- **Promote data communication through Semantic Web approaches**

What is the Semantic Web?

◆ The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. [W3C Consortium]

◆ The semantic web comprises a philosophy, collaborative working groups, and a variety of enabling technologies…all of which are intended to provide a formal description of concepts, terms and relationships within a given knowledge domain. [Wikipedia]

The Semantic Web and Allergens - What is Needed?

- **Terms and Relationships - A common vocabulary**
  - Allergen Ontology
- **Enabling Technology - Metadata**
  - XML
- **Collaborative Working Groups**
  - Data sharing and access

What is an Ontology?

- A data model that represents a set of concepts within a domain and the relationships between those concepts
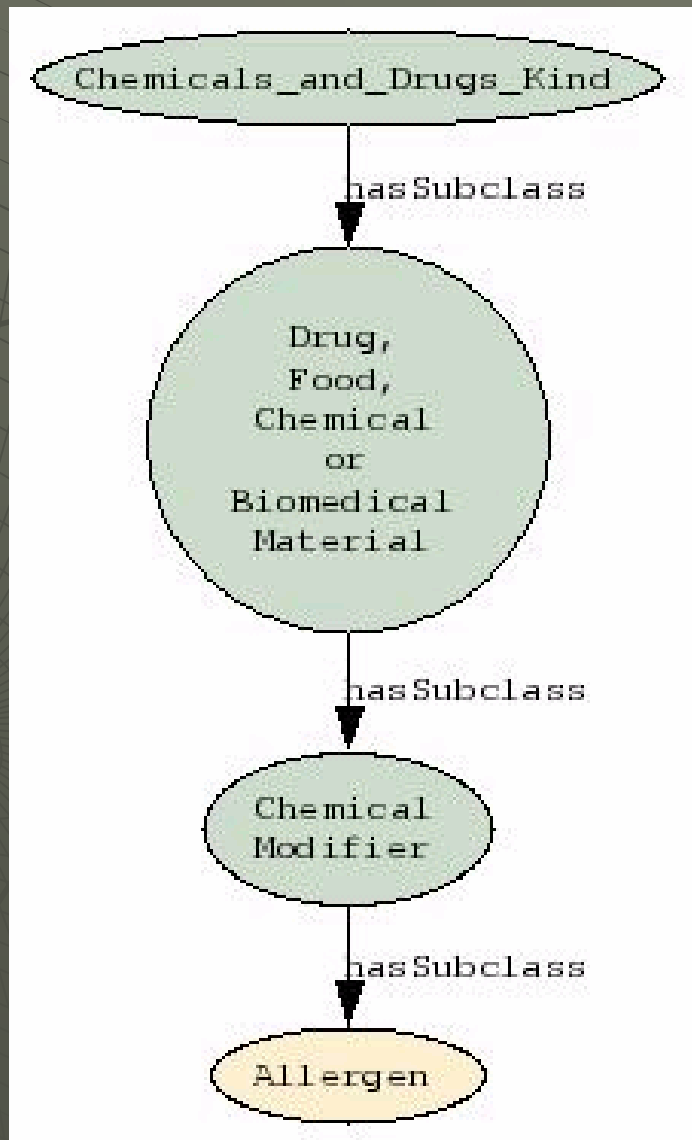
NCI Ontology - Concept Example

Class/Type Details

General

Class/Type Name	**Allergen**
Id	**C62651**

Attributes

Definition	**A substance that elicits an allergic reaction.**
PT	**Allergen**
primitive	**true**
CONCEPT_NAME	**Allergen**
Semantic_Type	**Chemical Viewed Functionally**
SYNONYM	**Allergen**

NCI Ontology - Relationship Example

NAL Ontology – Concept Example

allergens

Definition
Antigen-type substances that produce immediate hypersensitivity

Definition Source
MeSH (Medical Subject Headings)

Spanish
alérgenos

Broader Term
antigens

Related Term
allergenicity
hypersensitivity

NAL Ontology - Relationship Example

Legend
. = Narrower Terms
: = Broader Terms

[New Search] [Show Term Relationships]

allergens

: : : : biology
: : : immunology
: : immunologic factors
: antigens
allergens

How to Develop an Allergen Ontology

- ◆ **Create domain-specific terminology**
  - Example – allergen, major allergen, putative allergen, isoallergen, food allergen
- ◆ **Avoid ambiguity by using as many terms as necessary for all relevant concepts**
- ◆ **Include information on evidence**
  - Example - "Peptide epitope" or "digestion epitope" depending on the method used to identify the epitope
- ◆ **Start from and integrate with existing biomedical ontologies**
  - MeSH terms, NAL Thesaurus
- ◆ **Include relationship information**
  - Example: "is_a" to show membership in a larger category

The Semantic Web and Allergens - What is Needed?

- **Terms and Relationships - A common vocabulary**
  - Allergen Ontology
- **Enabling Technology - Metadata**
  - XML
- **Collaborative Working Groups**
  - Data sharing and access

What are Metadata?

- Metadata = data about data
- Already widely used – for example in structuring citation databases
- Can be used with a domain-specific ontology
- XML provides the format

Metadata – PubMed XML Example

```
<Journal>
    <ISSN IssnType="Print">1613-4125</ISSN>
    <JournalIssue CitedMedium="Print">
        <Volume>50</Volume>
        <Issue>7</Issue>
        <PubDate>
            <Year>2006</Year>
            <Month>Jul</Month>
        </PubDate>
    </JournalIssue>
    <Title>Molecular nutrition &amp; food research</Title>
    <ISOAbbreviation>Mol Nutr Food Res</ISOAbbreviation>
</Journal>
<ArticleTitle>Allergen sequence databases.</ArticleTitle>
<Pagination>
    <MedlinePgn>633-7</MedlinePgn>
```

The Semantic Web and Allergens - What is Needed?

- **Terms and Relationships - A common vocabulary**
  - Allergen Ontology
- **Enabling Technology - Metadata**
  - XML
- **Collaborative Working Groups**
  - Data sharing and access

What does Data Sharing and Accessibility Mean?

- Separating database content from search and analysis functions

- Providing a path to those data

Conclusion

- The tools for bioinformatic analysis of allergenic proteins continue to develop

- The utility of these tools will improve as more data and more types of data are generated (structures, epitope maps, etc.)

- Resources exist to allow data access and interoperability