HESI

Allergen and Toxin Databases

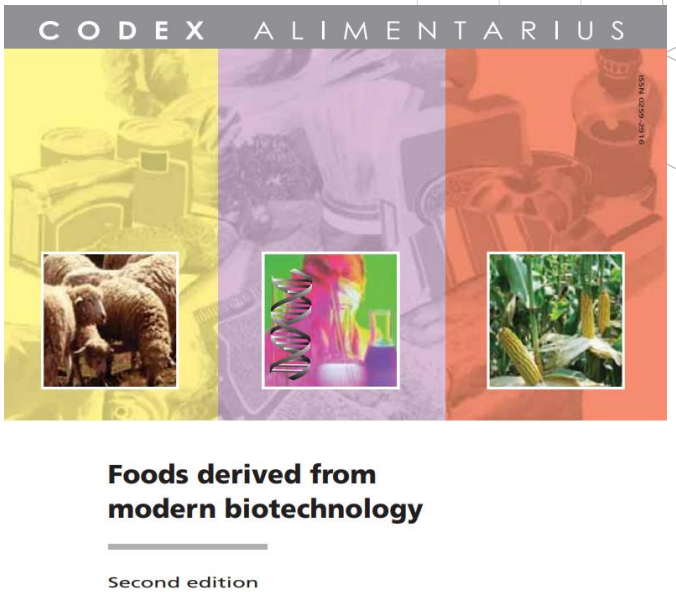
Andre Silvanovich/October 2020

Presentation Outline

- Regulations/Guidance
- Protein Allergenicity and Toxicity are Mechanistically Distinct
- Allergen Database
- Toxin Database
- Algorithms and Alignment Thresholds
- Alignment Interpretation
- When to use bioinformatic analysis
- Summary

Regulations/Guidance

- Allergy: CODEX 2003, 2009 and EFSA 2017
- Toxicity: CODEX 2003 & 2009 "the assessment of potential toxicity should focus on amino acid sequence similarity between the protein and known protein toxins and antinutrients (e.g. protease inhibitors, lectins) …"

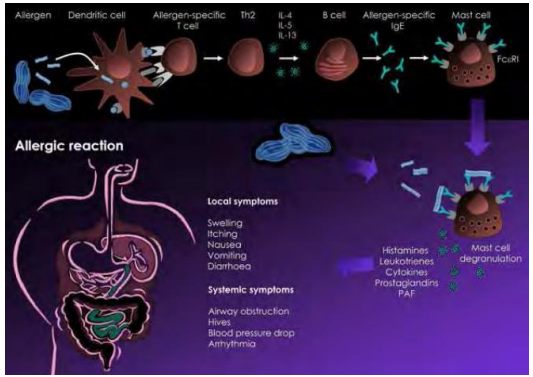


- Bioinformatic analysis of a protein sequence is useful for hypothesis formulation.
- Based upon the bioinformatics assessment:
  - Is it a good commercial candidate, no indication of toxicity
  - Is additional testing needed, if so, what?

HESI

Protein Allergenicity and Toxicity are Mechanistically Distinct

Parameter	Allergen		Toxin
Exposure to protein and outcome	Genetically predisposed individual	Sensitization	Acute toxicity in all members of a susceptible class
		Response	
Mechanism	Structure-IgE-binding sites, linear or discontinuous		Protein function
Cross reactivity	Yes		NA
Test of unknown protein	No definitive test for novel allergens		*In vivo or in vitro* cell-based assays



http://www.pgdc.ca/pdfs/plenary/2010/Janitha%20Wanasundara%20PGDC-2010-mustard%20allergens-Mar%203.pdf

https://www.fix.com/blog/foods-that-can-be-toxic/

HESI

Allergen Database

- Why has industry used an allergen database for >20 years?
  - Helps to focus bioinformatic search results
  - Not all allergens are identified as such
    - Typically protein discovery/identification precedes allergen designation
- How are allergens identified for inclusion in a database?
  - Sequence annotations in GenBank or Uniprot
  - IUIS (International Union of Immunological Societies) assignment
  - Publications
  - Annotation + publication + expertise to interpret
- No apparent relationship between allergenicity and function
  - Some relationship between structure and allergenicity
    - Cross reactivity
- No test for a novel sequence that is unrelated to known allergens

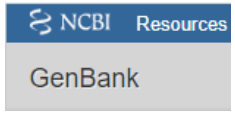
HESI

Allergen Database Continued…

- Industry supported efforts
  - HESI COMPARE database
  - UNLAllergenonline
- Scope of database
  - ~2200 unique sequences
  - ~220 Pfam members
  - Slow growth 1-3 novel pFams per year, most additions are homologues of existing entries
- Celiac peptide database

Toxin Database

- Toxin databases have seen less use by industry
  - The majority of “toxins” receive assignment by virtue of function
    - “The GO definition of “**toxin activity**”, i.e. proteins that interact selectively with one or more biological molecules in another organism (the "target" organism), initiating pathogenesis (leading to an abnormal, generally detrimental state) in the target organism. The activity should refer to an evolved function of the active gene product, i.e. one that was selected for.” EFSA 2020
- Scope of a “toxin” database
  - EFSA 2020 – 6,963 sequences
  - SwissProt – 9,575
  - GenBank 235 – 297,008 sequences (Genome assemblies/auto annotations)



- What considerations should be made when selecting toxins for a database?
  - Interaction with protein yields “abnormal, generally detrimental state”
  - Potential of specificity with “target organism”
  - Route of delivery and/or dosage
- Or does one use SwissProt or GenBank to identify function and assess alignments case-by-case?

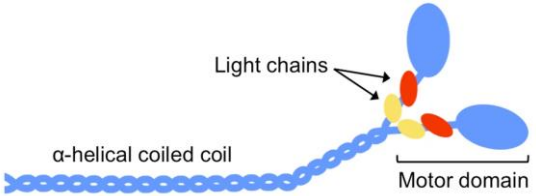
HESI

Algorithms and Alignment Thresholds

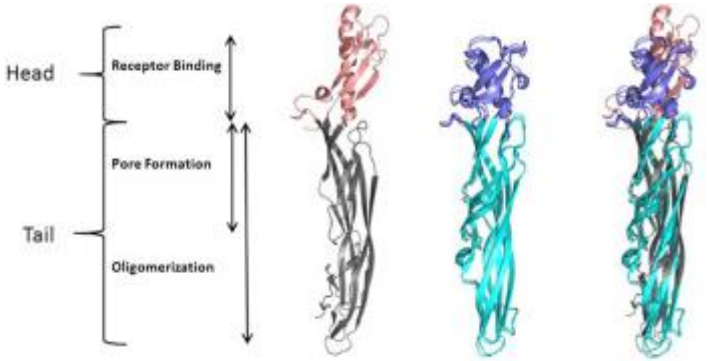
- Allergen
  - Sliding 8-mer window search
  - FASTA or BLASTp
  - Expectation value
  - 35% identity in 80 amino acids
  - Threshold exceeding alignments resolved case by case
- Celiac
  - Six identities in 9-mer
  - Threshold exceeding alignments resolved case by case
- Toxin
  - FASTA or BLASTp
  - Expectation value
  - Threshold exceeding alignments resolved case by case

Alignment Interpretation and Assessment

- Threshold exceeding alignments with proteins of concern do not necessarily indicate need for concern
- Safe proteins may share domains/motifs with proteins of concern
- Coiled-coil, a common structural motif found numerous proteins including safe proteins, and allergens and toxins
- Three domain β-pore forming proteins
  - Receptor-binding
  - Oligomerization
  - Pore forming
- In the absence of receptor-binding, oligomerization and pore formation don't occur
  - All three domains are required for the selected, evolved function
- Different receptor-binding domains impart different binding specificities



https://skeletalmusclejournal.biomedcentral.com/articles/10.1186/s13395-019-0192-3



https://www.sciencedirect.com/science/article/pii/S0273230017302167?via%3Dihub

HESI

When to use Bioinformatic Analysis to assess Toxicity?

- As early as is feasible in the product development pipeline
  - Drives go/no go decisions prior to investment of effort
- Repeat regularly during product development up to the point of toxicity testing
- Based upon weight of evidence, once a conclusion of safety is made, additional bioinformatic analysis is unlikely to further inform

Summary

- While the purpose of bioinformatic assessment of allergenicity and toxicity is identical, mechanistic differences afford different approaches
- Allergenicity assessments benefit from a database that contains curated sequences
  - Protein function is independent of allergenicity
  - With bioinformatic insight to drive a hypothesis, it is possible to test for cross reactivity
  - No test can predict novel allergenicity
- In contrast, toxicity assessments are less likely to benefit from a database that contains curated sequences
  - Protein toxicity is related to protein function, searches of UniProt/SwissProt or GenBank are sufficient to make functional identification
  - Domain-function identification in alignments is useful hypothesis formulation
  - Definitive tests for protein toxicity are available

Thank you