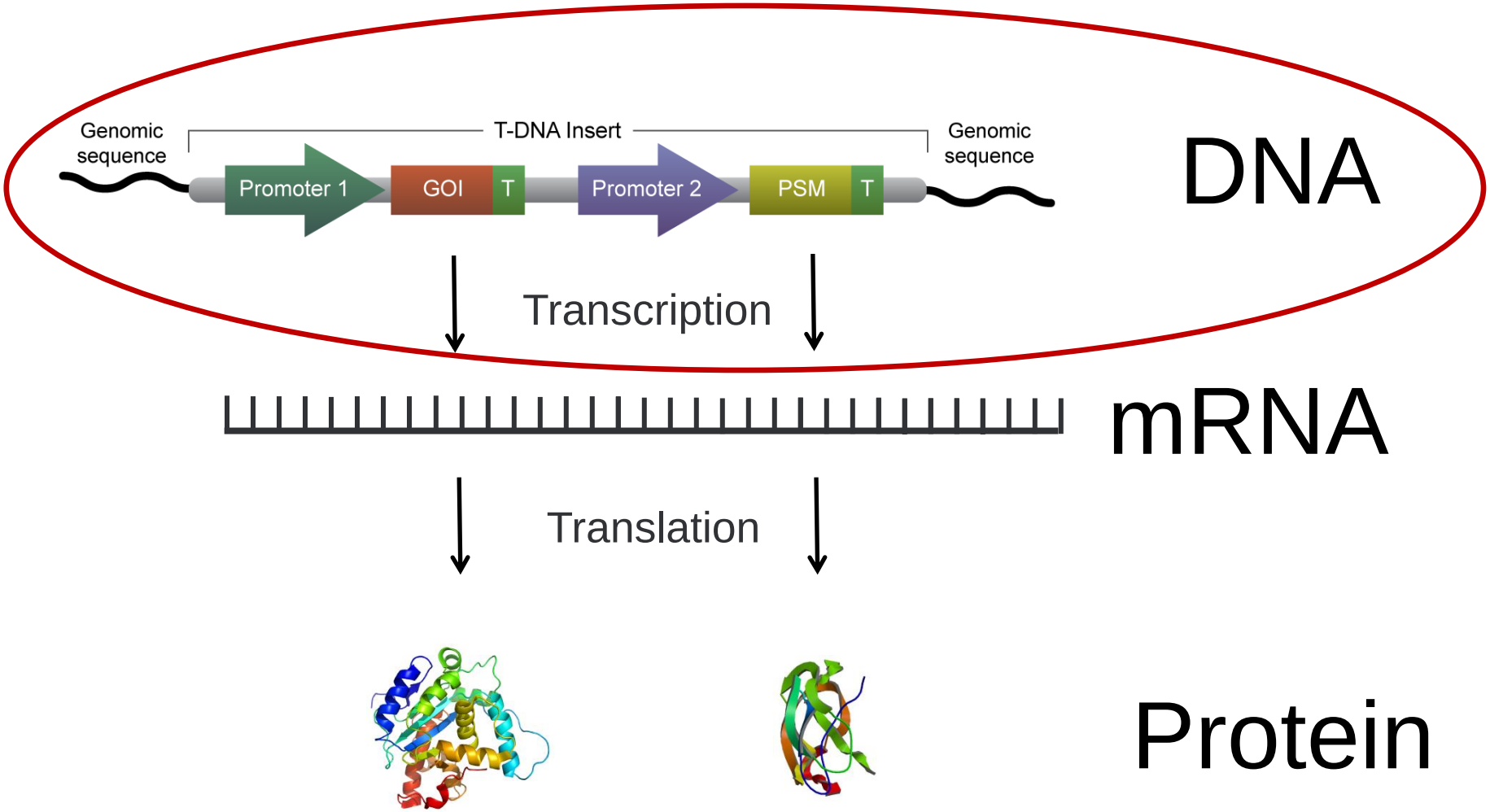
Molecular and Protein Characterization of GMO Products

Beijing, China 15-16 April 2013

Scott McClain, Ph.D.
Syngenta Crop Protection, LLC
Research Triangle Park, NC, USA

Molecular characterization



Molecular characterization is a key to communicating the properties of inserted DNA in a genetically modified plant

The inserted molecular component (DNA) of a transgenic crop consists of the gene(s) supporting expression of a protein(s) with specific trait(s) and supporting DNA, such as promoters and terminators.

T-DNA = transferred and inserted DNA that expresses one or more of the intended proteins that provide the insect resistance or other trait

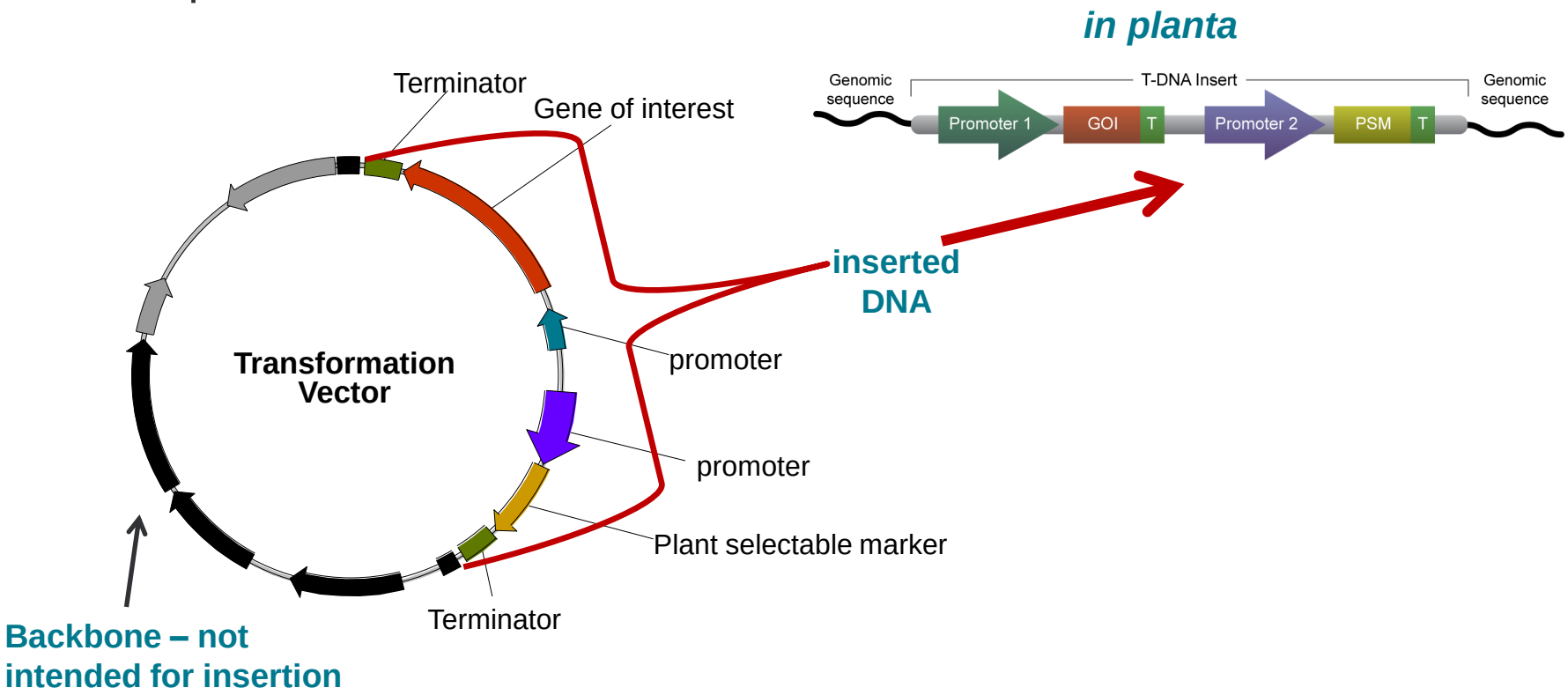
GOI = gene of interest

Molecular characterization answers important questions about the inserted DNA

- What DNA was put into the crop?
- How many genes were put into the crop?
- Where in the host genome is the inserted DNA located?
- Is expression of the gene(s) stable?

What DNA was put into host crop?

- Vector Sequence – intended for insertion through the transformation process
- Product (GMO Plant) Sequence Analysis – what was actually inserted into the plant



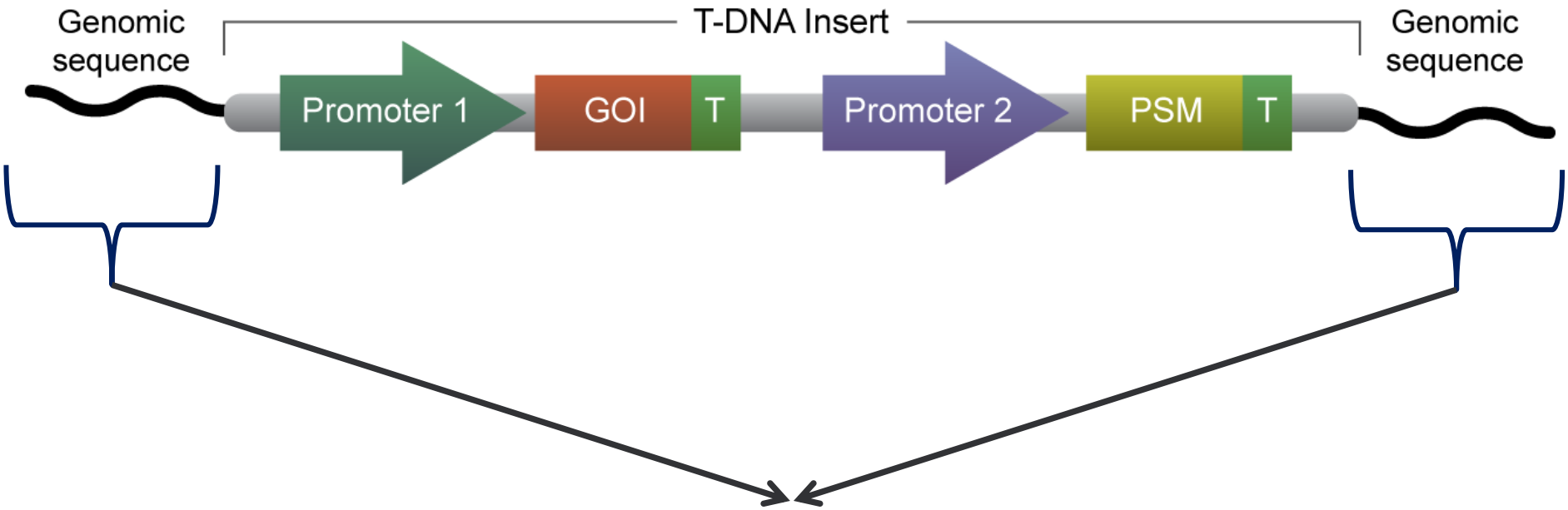
How many genes were put into the crop?

- Insert Sequence Analysis; DNA from plant is evaluated by sequencing analysis
- Copy Number is evaluated by Southern Analysis

Where was it put into the crop's genome?

Genomic Sequence Analysis
Was a crop gene interrupted? – scan the genomic sequence on either side and then perform bioinformatic search to determine nearby genes, if any.
At the site of insertion, were any changes made to the native genomic sequence?
On what chromosome is this transgenic insert located?

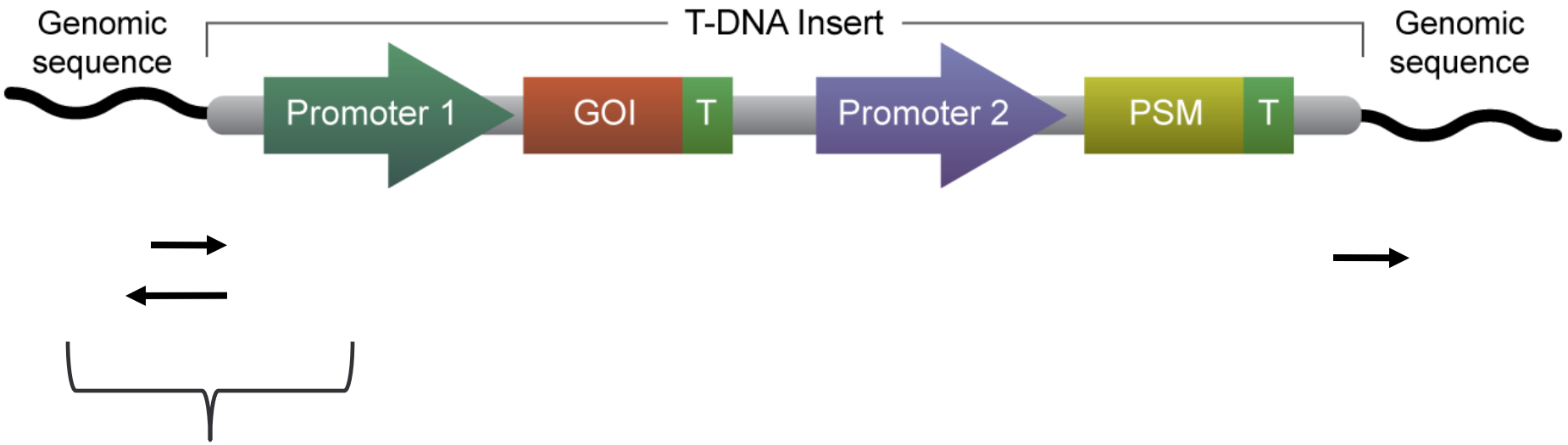
The location of the insert DNA is critical to understanding the location of nearby genes, if any, in the host genome



Once location of the insert DNA is confirmed, sequence of the nearby “flanking” DNA from the host plant is used to bioinformatically identify genes, if present. Sequence is compared to the genome libraries of the species of the host.

Is there potential to produce any unintended proteins?

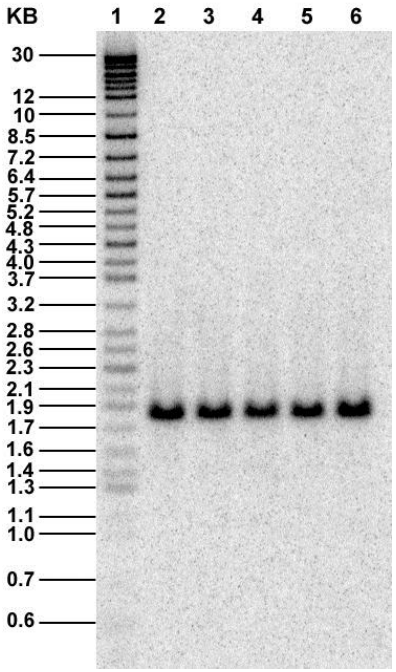
Determine if there are any sequences that could potentially result in an unintended protein. A putative fusion protein created between the insert and genomic DNA would be unintended (and unlikely).



Evaluate whether sequences with stop codons on either side of genomic/insert junction are present and determine if they are significantly similar to allergens or toxins – bioinformatics is used to support this.

Has the insert DNA been transferred to the host plant in a stable manner?

Is the insert DNA still present through successive breeding generations?



Lane 2 : Generation T3

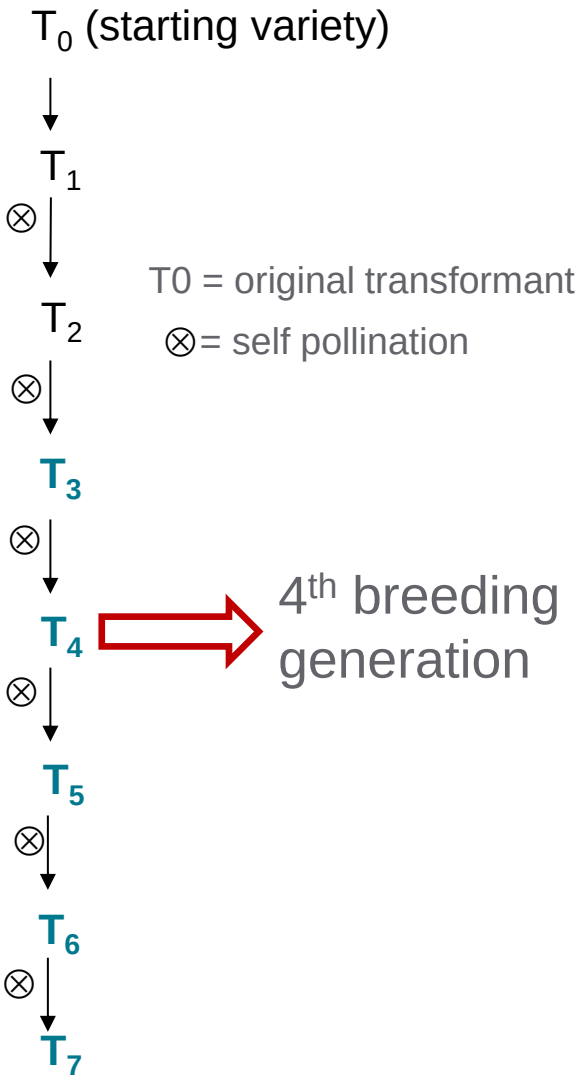
Lane 3: Generation T4

Lane 4: Generation T5

Lane 5: Generation T6

Lane 6: Generation T7

Genetic Stability - Southern DNA Analysis



Additional check for molecular stability of the insert DNA

Mendelian Inheritance

- Does the insert inherit according to Mendel's Law of Segregation?
- For example, when a heterozygous parent is crossed with a heterozygous parent, do the offspring have a positive to negative ratio of 3:1?

Heterozygous Transgenic Parents

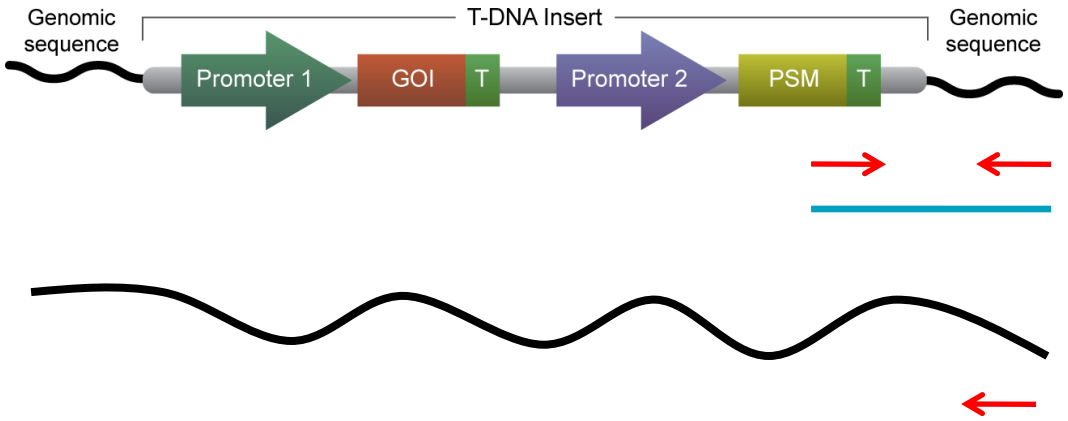
	+	-
+	+/ -	+ /-
-	+/ -	- / -

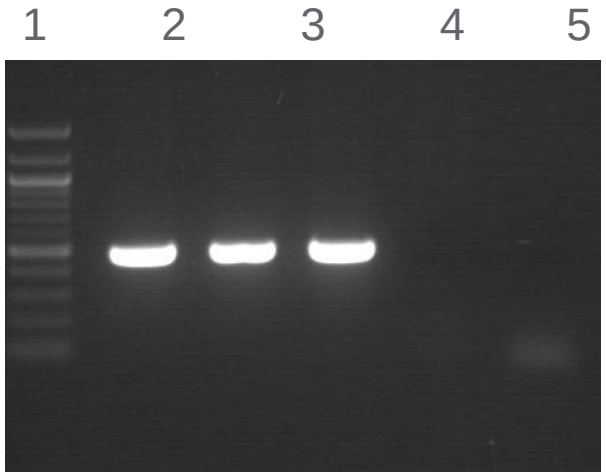
Molecular detection methods

- After genetic trait stability and safety have been determined for the transgenic crop, the following questions remain - What GMO product is in the market?, How many different kinds?, Where are they located?
- In order for the transgenic crop to be registered there must be a method for detecting this transgenic crop.
- Example: detection methods for grain are typically required as part of the regulatory dossier in many parts of the world.

Expected detection methods

- Event-specific ***Qualitative*** PCR method (presence or absence)
- Event-specific ***Quantitative*** PCR method (quantity)





Sample 1 – GM

Sample 2 – GM

Sample 3 – GM

Sample 4 – non GM

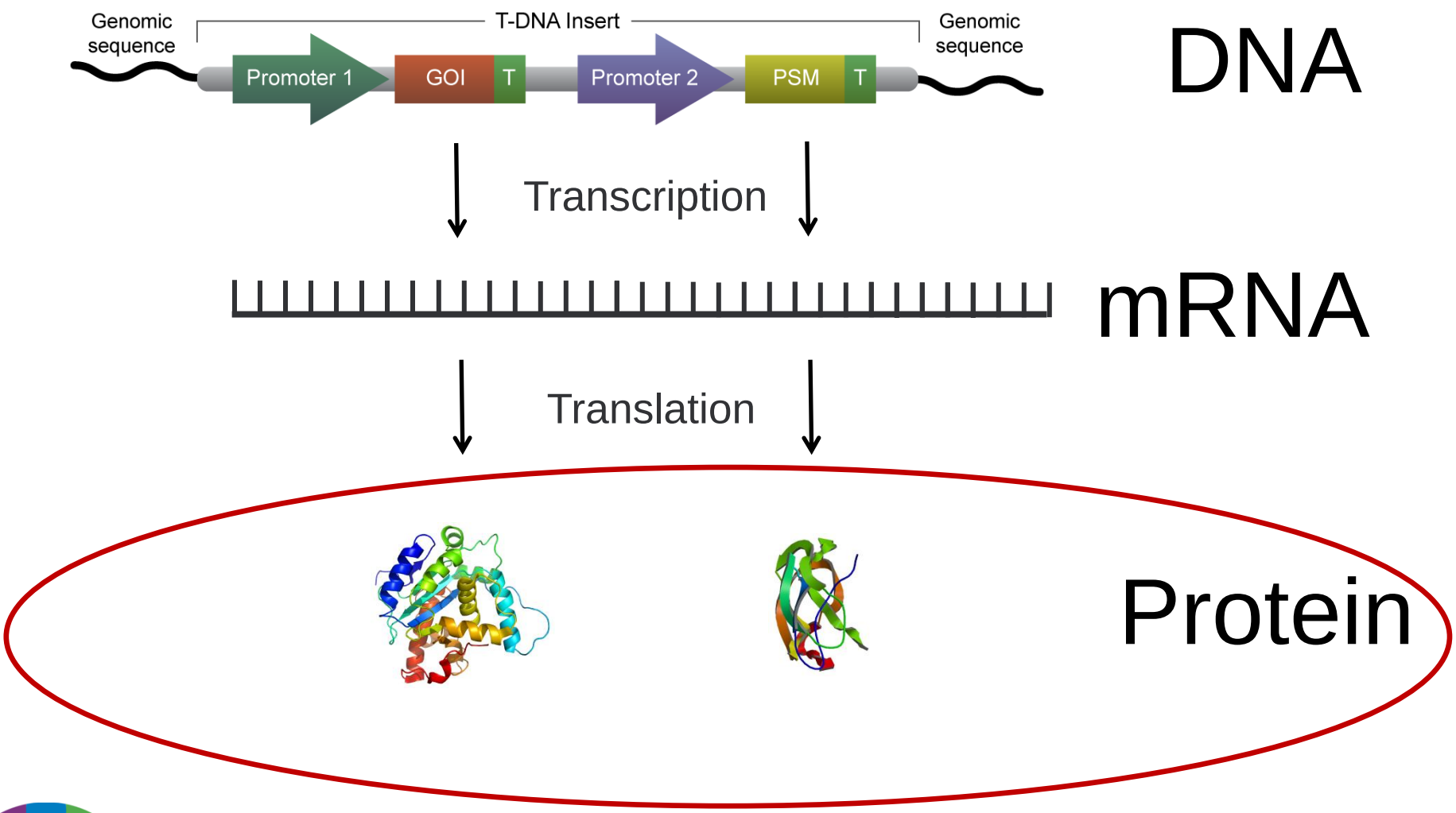
Sample 5 – non GM

Molecular Characterization - Summary

Provides confirmation to regulators of GMO crops that the inserted DNA is:

- Fully described and exactly as intended when transferred to new crop host
- Is known for location in the host genome and lack of interference with endogenous genes
- Is stable from a breeding perspective
- Expresses intended genes
- Can be detected when placed into commercial market

Protein characterization



Protein Characterization

Protein Characterization studies are performed to characterize the transgenic GMO protein(s). They address several questions, including:

1. Is the transgenic protein expressed in the plant in a stable manner?
2. Are the biophysical properties of the transgenic protein in the plant consistent with a safe protein?
    a. Do the protein properties support food, feed and environmental safety?
3. Is the E. coli-produced protein biochemically and functionally equivalent to the GMO plant protein?
    a. Can recombinant GMO protein be used for safety studies?
4. How does the protein behave in different solutions used in toxicity and eco-toxicity studies?

Protein Characterization – analytical approaches

Biochemistry is used to assess physical properties

- SDS-PAGE and Western blot (Molecular weight)
- activity assays (for enzymes)
- total protein quantitation
- densitometry
- % purity calculations for trait protein
- mass spectrometry (intact mass analysis)
- N-terminal amino acid sequence analysis

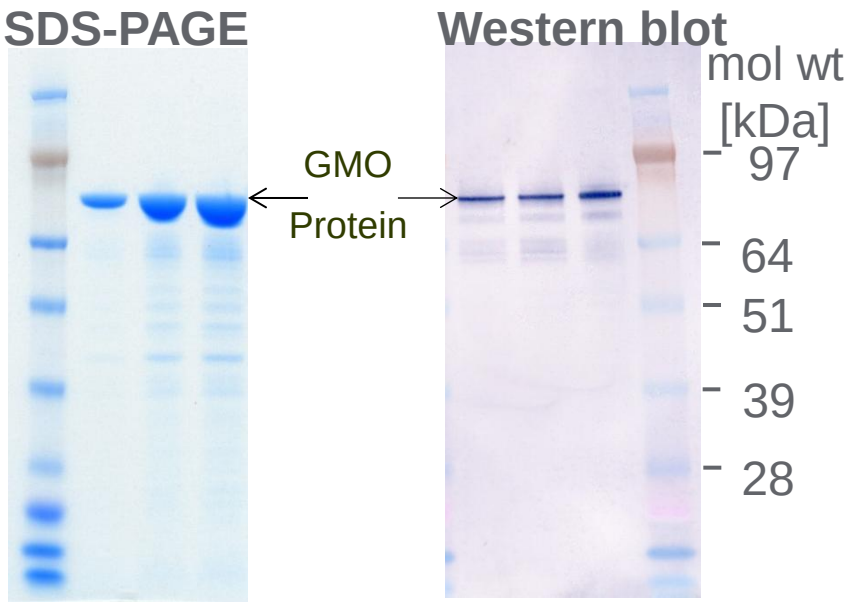
- **Bioinformatics** – Comparison of amino acid sequence from the transgenic crop to those of the non-transgenic crop as well as sequences from other organisms.

Characterizing the *E. coli*-produced GMO protein

- Prior to using in any regulatory studies, the purified protein is characterized according to Good Laboratory Practices

Test Substance Characterization
Can the identity of the protein in the test substance be confirmed?
What buffers can the test substance be dissolved in?
What is the purity of the test substance?
Is the protein active after purification from *E. coli*?

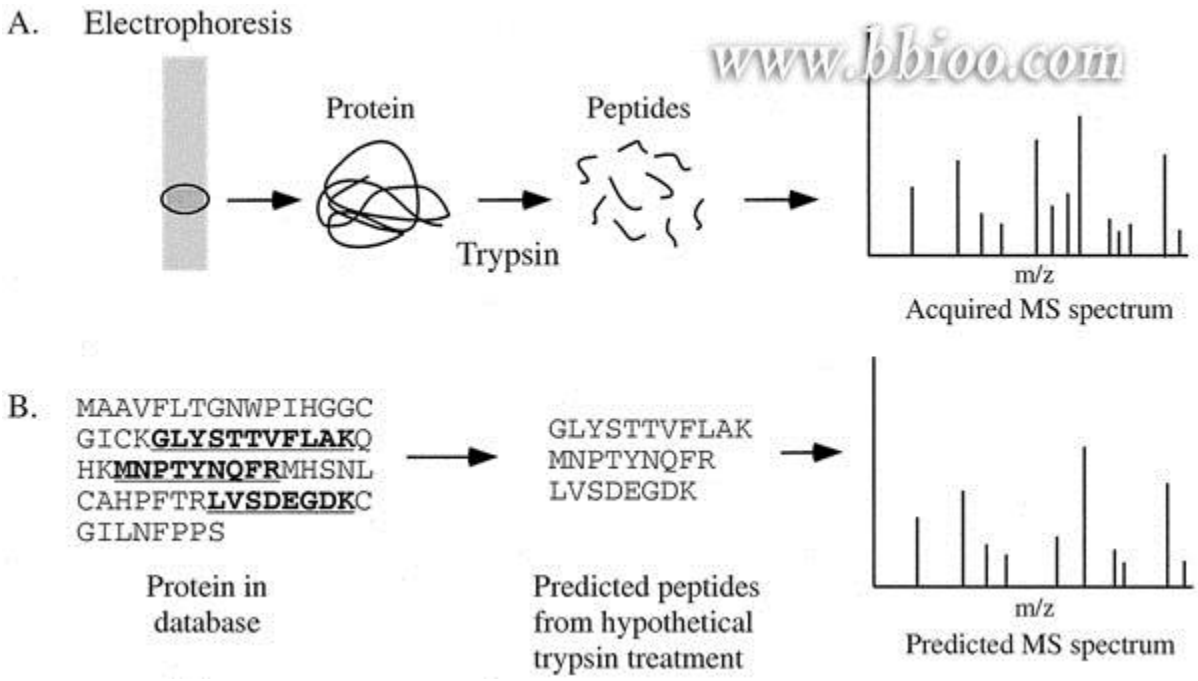
- ✓ SDS-PAGE
- ✓ Western blot
- ✓ Activity assay



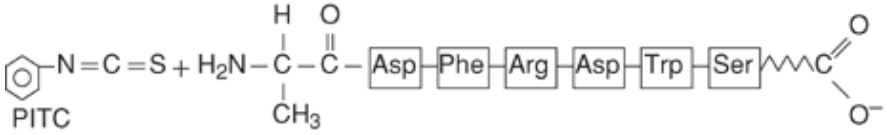
Protein mass analysis and N-terminal sequencing analysis

- Intact mass analysis: Q-TOF, MALDI-TOF

- Peptide mass mapping:



- N-terminal sequence analysis (Edman degradation):



Protein characterization, continued

Heat stability study	✓ ELISA ✓ Activity Assay
Is the protein stable after treatment at different temperatures?	

pH stability study	✓ Activity Assay
What is pH optimum of the expected reaction? This is only performed on enzymes.	

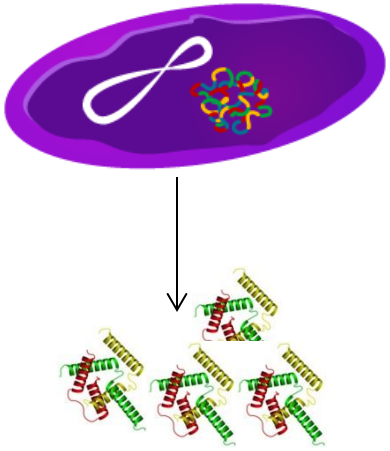
Substrate specificity study	✓ Activity Assay
Does the enzyme have high or low substrate specificity?	

GMO protein production for safety testing

Transgenic Plant

The concentration of the expressed transgenic protein in the plant tissue is low

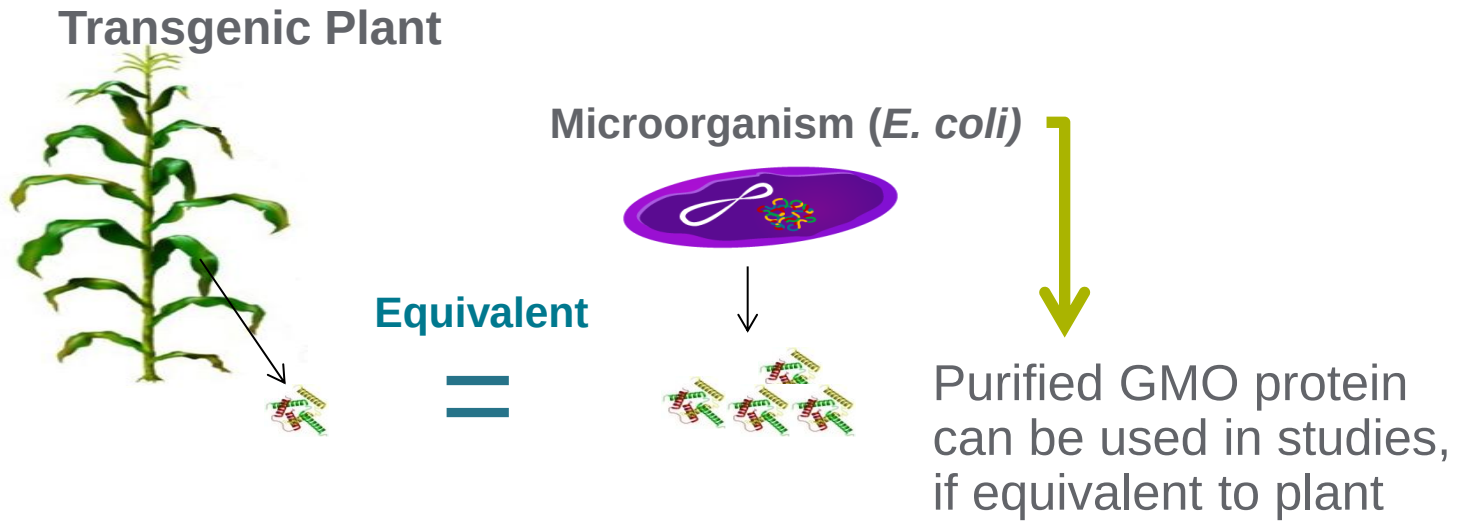
E. Coli host for production of purified GMO protein



A microbial host is used to make large quantities (50 grams or more) of the GMO protein

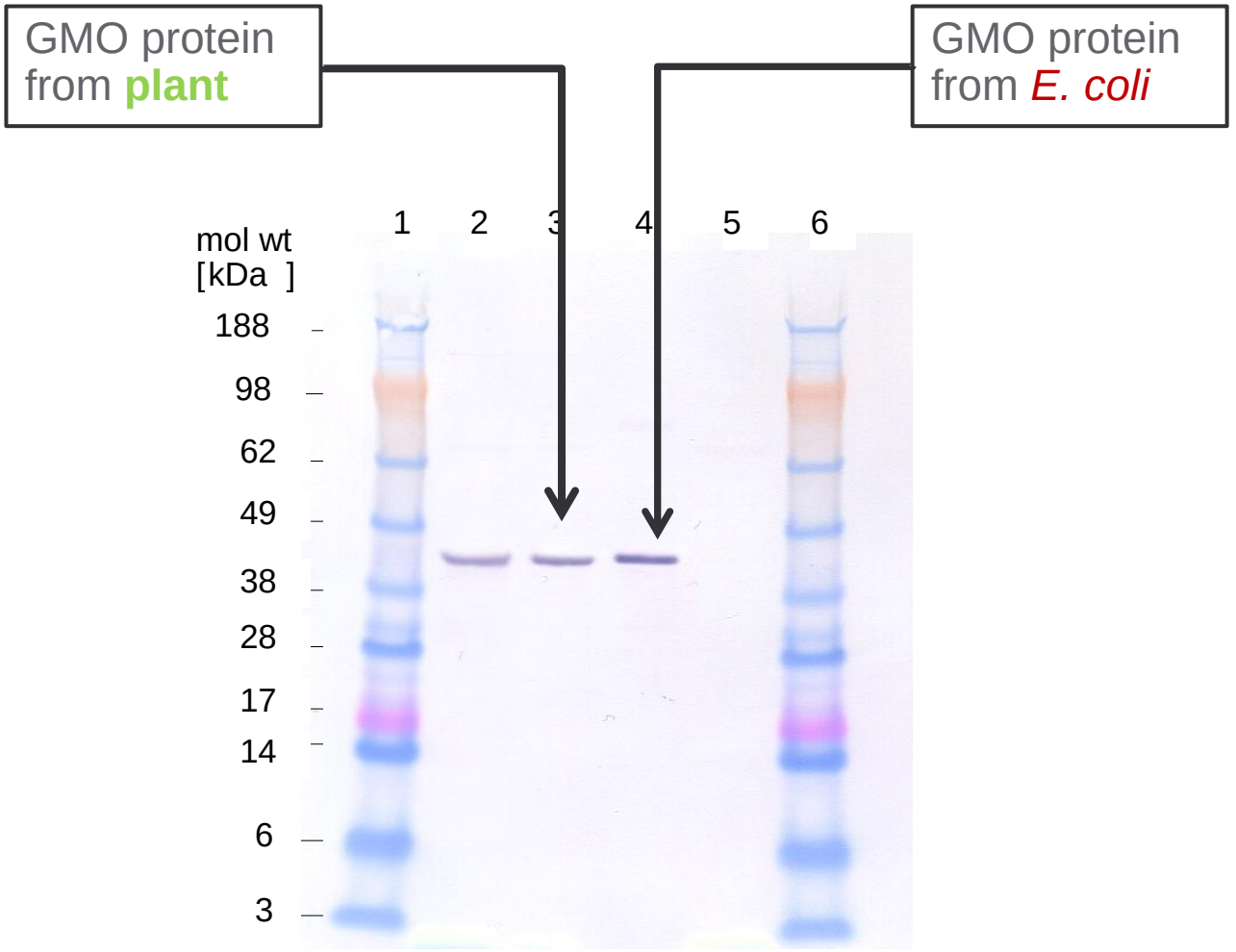
Microbial produced protein is a surrogate for plant expressed transgenic protein in safety studies that require large quantities of purified protein.

Is the microbial (*E. coli*) protein equivalent to the plant expressed protein?



Equivalence (Bridging) Study
Do both protein have the same predicted molecular weights?
Do both proteins react with the same antibodies?
Are both proteins not glycosylated?
Do both proteins have the same identity?
Are both proteins active and have comparable activities?

Protein from both sources is expected to be same molecular weight and react with same antibody in Western blot



SDS-PAGE and Western blot using protein-specific antibodies

Protein characterization studies support specific safety studies and the allergy assessment

- Characterized *E. coli*-produced GMO protein is used for;
  - Characterization studies to ***assess allergenic potential***
    - SGF
    - SIF
    - Glycosylation
  - Acute toxicity study in mice
  - Animal studies for environmental exposure assessment
  - Additional animal studies that may be requested

Allergy assessment for biotech foods: characterization of GMO protein for lack of allergen similarity

- Categories of public health concern regarding potential protein allergenicity and biotechnology:
  - Transfer of **known** allergens or **cross-reactive** proteins
  - ***Novel proteins*** *becoming allergens in sensitive populations*
  - *Increasing* ***endogenous*** *allergenicity of an already allergenic crop*

Allergy assessment, continued

- **Purpose:** Determine if the biotech protein is a known allergen or cross reactive with a known allergen.
  - No single, definitive test of protein allergenicity exists
  - ‘Weight of evidence’ approach recommended by Codex Alimentarius guidelines is used to determine the allergenic potential of the biotech protein
  - Weight-of-evidence approach means several characterization studies are performed to assess the overall likelihood of allergenic potential.
    - Is the protein from an allergenic source?
    - Bioinformatics assessment of biotech protein
    - IgE-binding study using serum from an allergic patient, if appropriate

Bioinformatics: Source of the GMO protein

- Is the protein from an allergenic source?
  - Allergenic sources are defined by Codex Alimentarius as those that have reasonable evidence of IgE-mediated oral, respiratory or contact allergy. Also defined by proteins listed in University of Nebraska www.allergenonline.org.
  - If protein originates from an allergenic source, Codex recommends performing a serum study using serum from appropriately allergenic individuals
  - Serum from allergic individuals that are reactive to the source of the protein would be required, if necessary.
    - Serum contains IgE antibodies to allergenic source proteins
    - Serum study will determine if there is IgE reactivity to the biotech protein

Bioinformatics: Assess sequence similarity to known allergens

- Purpose: Compare the GMO protein to a database of known allergens.
  - Allergen database (allergenonline.org) is updated annually and curated by international academic allergy experts
  - Criteria for inclusion of a protein in the allergen database
    - Protein binds IgE from individuals with clear evidence of allergy to the source of the protein
    - Protein causes basophil activation or histamine release, skin test reactivity or challenge test reactivity using subjects allergic to the source
- Types of comparisons
  - Whole sequence homology (FASTA analysis)
    - Sliding window (80-mer) search has been performed in past
  - Small peptide (8 amino acids) matching

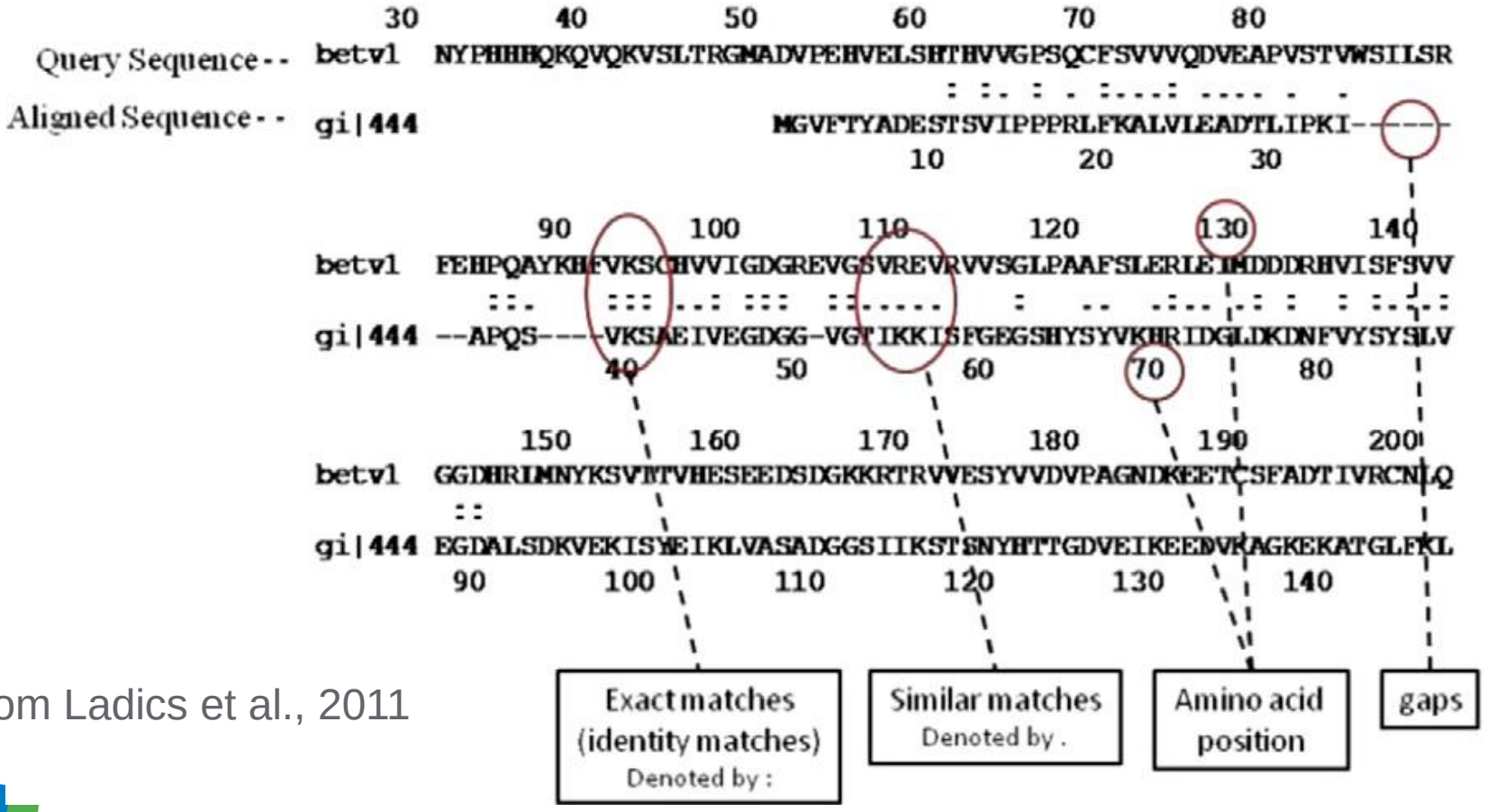
Bioinformatics, continued

- Full length FASTA protein alignment is used to compare GMO protein to each allergen in the database

- **Goal:** Determine if the biotech protein is an allergen or may be cross-reactive with a known allergen by determining shared structure
    - When does a protein have significant sequence similarity to an allergen?
        - **≥35% shared identity over 80 or greater amino acids**
            - Based on Bet v 1 cross reactivity (Aalberse et al., 2001)
        - E-score cut-off can be used to be as conservative as ≥35% identity over 80 amino acids criteria
            - Based on work by Silvanovich et al., 2009

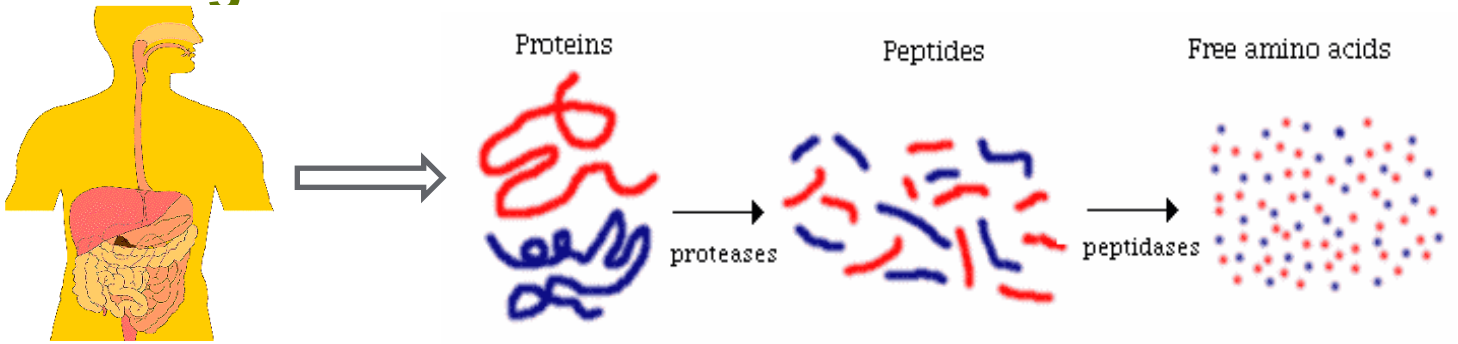
Example of bioinformatic alignment with allergen database

Description: GI-|44409474|gb|AAS47036.1| *major cherry allergen Pru a* (160 aa)
FASTA parameters: initn: 44 init1: 44 opt: 116 Z-score: 155.0 bits: 35.3 Smith-Waterman score: 119;
Identity and Similarity: 27.95% identity (56.989% similar) in 93 aa overlap, E-Score = 0.001



From Ladics et al., 2011

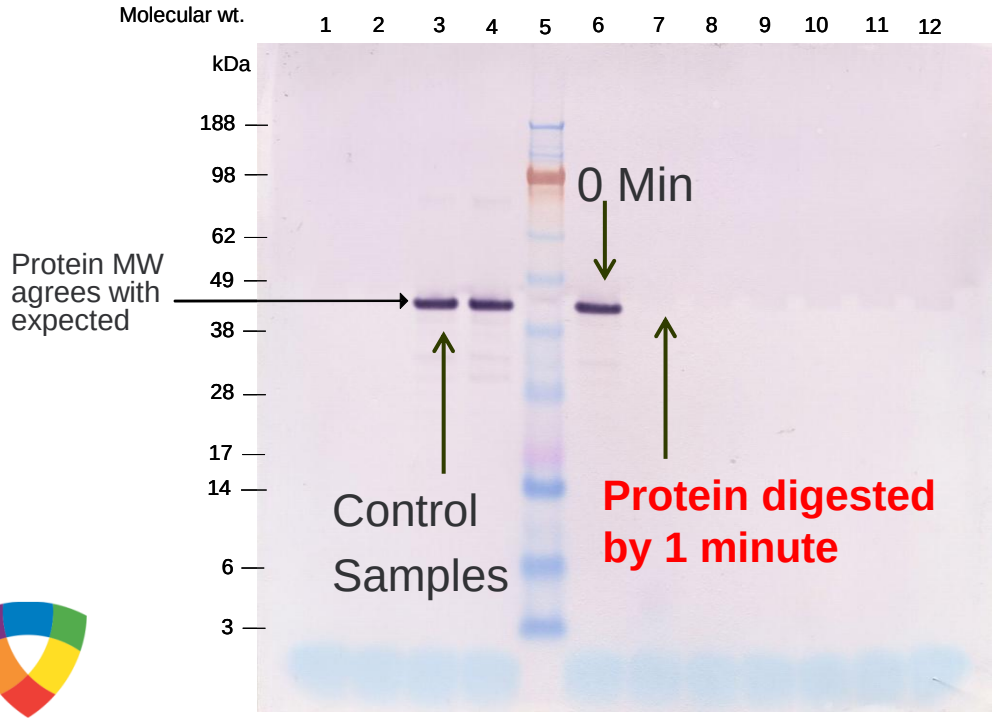
Pepsin Digestibility - Does the protein behave like known allergens?



Some known allergens are stable to pepsin digestion

In vitro digestibility study

Is the protein stable in ***simulated gastric fluid*** (SGF) or ***simulated intestinal fluid*** (SIF)?

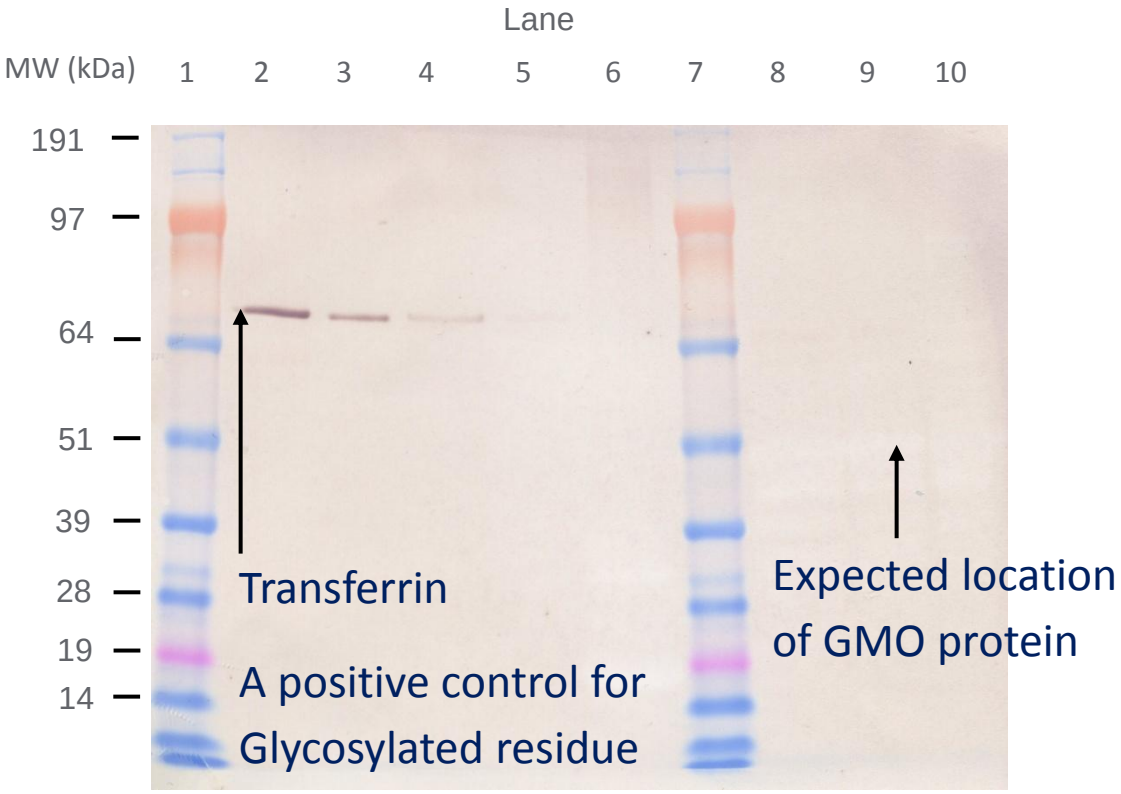


Quickly digested proteins indicate low likelihood of gastric exposure to immune system

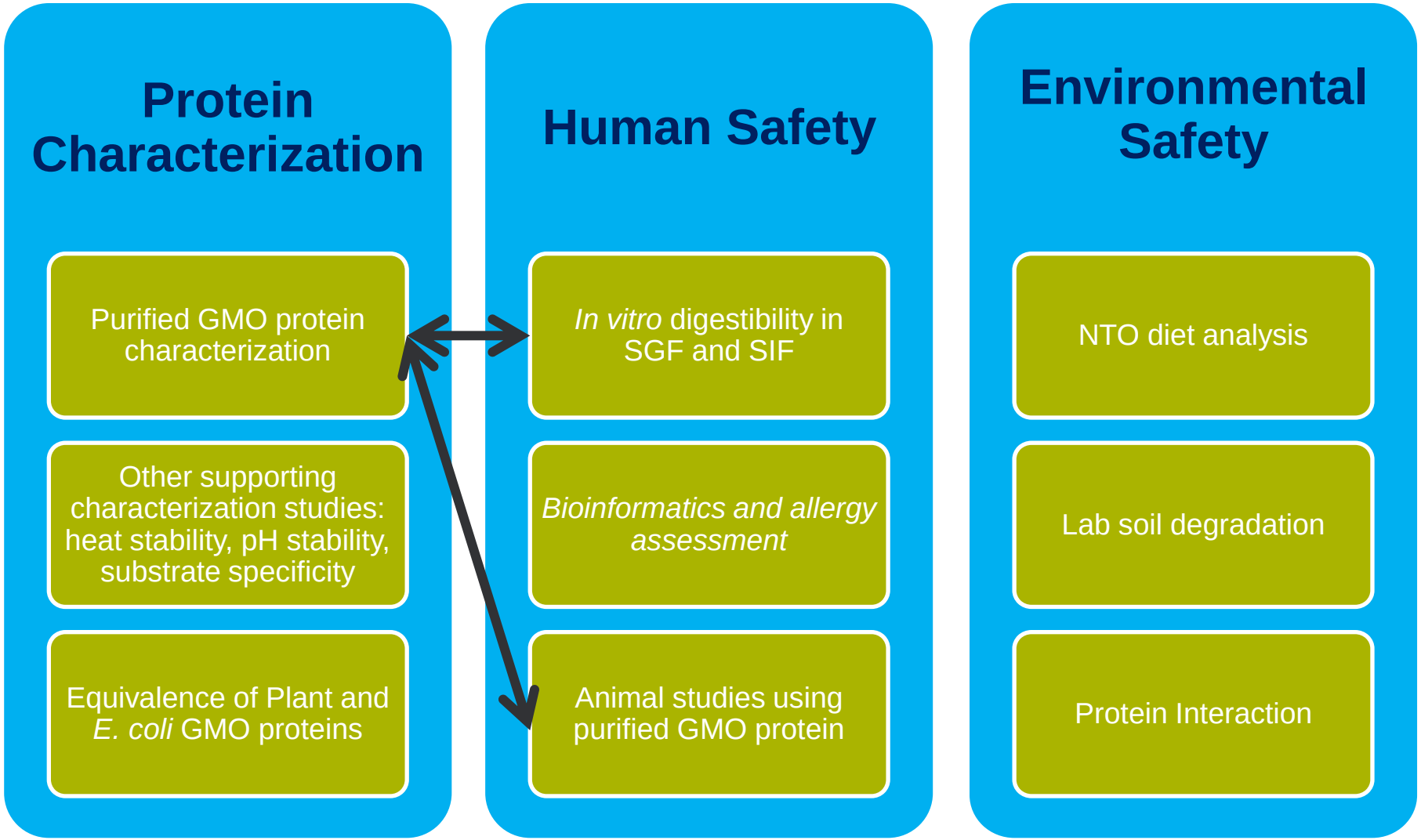
Glycosylation – Is the plant-produced GMO protein modified?

SDS-PAGE is used to separate proteins by molecular weight and reagents specific to glycosylated residues are used for detection.

Plant-expressed GMO proteins are expected to not have post-translational changes.



Summary – Protein characterization and safety assessment



Thank You

HESI-PATC bioinformatics publications

- Thomas K, Bannon G, Hefle S, Herouet C, Holsapple M, Ladics G, MacIntosh S, Privalle L. (2005). In silico methods for evaluating human allergenicity to novel proteins: International Bioinformatics Workshop Meeting Report, 23-24 February 2005. Toxicol Sci. 88(2):307-10.
- Ladics GS, Cressman RF, Herouet-Guicheney C, Herman RA, Privalle L, Song P, Ward JM, McClain S. (2011). Bioinformatics and the allergy assessment of agricultural biotechnology products: industry practices and recommendations. Regul Toxicol Pharmacol. 60(1):46-53.

- Other publications referenced in this presentation

  - Aalberse RC, Akkerdaas J, van Ree R. (2001). Cross-reactivity of IgE antibodies to allergens. Allergy. 56(6):478-90.
  - Ladics GS, Bardina L, Cressman RF, Mattsson JL, Sampson HA. (2006). Lack of cross-reactivity between the Bacillus thuringiensis derived protein Cry1F in maize grain and dust mite Der p7 protein with human sera positive for Der p7-IgE. Regul Toxicol Pharmacol. 44(2):136-43.
  - Silvanovich A, Bannon G, McClain S. (2009). The use of E-scores to determine the quality of protein alignments. Regul Toxicol Pharmacol. 54(3 Suppl):S26-31.