Considerations for safety assessment of intractable proteins expressed in genetically modified crops

Rong Wang*[a,1] John Zhang*[b], Matthew L. Cheever[c], Tao Geng[a], Rod Herman[d], Matthew G. Pence[e], Barry W. Schafer[f], Brandy Verhalen[b], Helene Vigeolas[g], Yanfei Wang[a], Jianhong Wu[h], Pei-Ying Wu[c], Christopher Ament[i],

[a]Bayer CropScience, 700 Chesterfield BLVD West, Chesterfield, MO, United States

[b]Corteva Agriscience, 8325 NW 62nd Ave, Johnston, IA, United States

[c]BASF Corporation, 26 Davis Drive, Research Triangle Park, NC, United States

[d]Corteva Agriscience, 9330 Zionsville Road, Indianapolis, IN 47968, United States

[e]J.R. Simplot Company, Simplot Plant Sciences, 5369 West Irving Street, Boise, ID, United States

[f]Schafer Scientific Solutions, 880 Monon Green Blvd, Suite 101, Carmel IN, United States

[g]BASF Belgium, Technologiepark-Zwijnaarde 101, 9052 Gent, Belgium

[h]Syngenta, 9 Davis Drive, Research Triangle Park, NC, United States

[i]Eurofins Food Chemistry Testing Madison, Inc., 6304 Ronald Reagan Ave., Madison, WI, United States

*These authors contributed equally to this paper as co-first authors.

ABSTRACT

Current requirements for safety assessment by regulatory agencies necessitate the production of multi-gram quantities of functionally active proteins newly expressed in genetically modified (GM) crops. This presents challenges for so-called intractable proteins (those that are difficult to isolate in an active form). Significant experience and knowledge have been gained by both GM crop developers and global regulatory agencies regarding the safety assessment of the newly expressed proteins (NEPs). By leveraging this knowledge and scientific data that have accumulated over the last several decades, the safety assessments of NEPs can be refined and improved. In this technical review, we examine the options for characterizing intractable NEPs and producing protein test substances from different sources to support the safety assessments of novel GM crops.

Keywords: weight of evidence, intractable proteins, newly expressed proteins (NEPs), safety assessments, risk assessment, genetically modified (GM) crops, protein isolation, protein characterization, history of safe use (HOSU), allergenicity, toxicity.

1 Introduction

Most genetically modified (GM) crops that have been developed and commercialized are widely grown and globally traded (Klümper & Qaim, 2014). GM crops are subjected to a rigorous regulatory approval process in both cultivation and importing countries. Legal

[1] Corresponding Author: Rong Wang. Email: rong.wang@bayer.com

regulatory requirements and industry stewardship are in place to ensure safety, responsible management and use of the technology and GM crop products before, during, and after commercialization (European Commission, 2010; Mbabazi et al., 2021; McHughen & Smyth, 2008). While proteins are necessary components of diets and the vast majority of them are safe for consumption by humans and animals, a small number of dietary proteins are known to possess toxic, antinutrient, or allergenic properties. Nevertheless, all newly expressed proteins (NEPs) introduced into GM crops are thoroughly evaluated for human and animal safety (Delaney et al., 2008; Koch et al., 2015; National Academies of Sciences, 2016; Nicolia et al., 2014; Prado et al., 2014).

1.1 Current Paradigm for Protein Safety Assessment

Protein safety assessments for food and feed use primarily focus on the potential toxicity and allergenicity of the NEP. Prior to safety assessment, NEPs are characterized to determine if the protein expressed in the GM crop has the expected biochemical properties and function. The low expression of NEPs in plants (*e.g.,* a milligram NEP in a kilogram grain) requires over-expression using heterologous (*e.g.,* microbial) expression systems to produce enough surrogate test substances for safety evaluation. Plant-produced NEPs are also compared with heterologously produced protein test substances to demonstrate that the test substance is a suitable surrogate used for laboratory-based studies (Brune et al., 2021; Raybould et al., 2013). The current safety assessment standards (Codex Alimentarius Commision, 2003, 2009; OECD, 1993; Sharma et al., 2022) were developed based on the scientific knowledge available at the time. The current process has several components, typically including: 1) reviewing of the history of safe use (HOSU) of the protein, its source organism, and the NEP's mode of action; 2) *in silico* bioinformatic comparisons of the NEP amino acid sequence with that of known toxins and allergens; 3) *in vitro* evaluation of the thermal stability of the NEP for use in food and feed, and stability during simulated gastric and intestinal digestion; 4) the NEP expression level in relevant plant tissues; and 5) performing *in vivo* studies with animals (typically using mice as the test system) at a hazard limit-dose. The routine *in vivo* toxicity studies which are often required even when no potential toxicity hazard is identified, require a large quantity of purified active proteins. For cultivation approvals, an environmental risk assessment is conducted, which includes assessment of potential exposure and/or hazard to non-target organisms (NTOs). If the NEP has an insecticidal mechanism, additional NTO studies may be conducted using purified proteins.

1.2 Intractable Proteins and Regulatory Challenges

Some proteins have been described by Bushey et al. (2014) as intractable proteins with characteristics including: 1) inability to express at high levels in either the crop or heterologous systems; 2) low solubility; 3) instability during or after extraction; and 4) inability to isolate an active form. These properties make it extremely difficult or impossible to express the intractable proteins in heterologous systems; to isolate, purify, concentrate, or quantify the proteins due to low levels; to demonstrate biological activity; or to prove equivalency with proteins expressed in plants. Five categories of intractable proteins have been identified (Bushey et al., 2014): 1) membrane-bound proteins; 2) signaling proteins; 3) transcription factors; 4) N-glycosylated proteins; and 5) resistance proteins (R-proteins) .

The characteristics of intractable proteins warrant additional considerations for conduct of the safety assessment. Some regulatory agencies currently request an acute oral toxicity study while European Food Safety Authority (EFSA) requires a 28-day repeated-dose toxicity study. These studies need up to 100 grams of purified active protein. It has been questioned whether the default regulatory requirements for high-dose rodent studies in assessing the safety of NEPs in GM crops are scientifically warranted. Given the difficulties in isolating multiple-gram quantities of active intractable proteins, the current regulatory paradigm for characterizing and assessing the safety of the proteins is unattainable. In addition to a more science-based approach for safety assessment, considerations of test materials and methods are necessary to evaluate the same endpoints currently used to assess the safety (Colgrave et al., 2019; Habig et al., 2018; MacIntosh et al., 2021; Madduri et al., 2012; Skinner et al., 2016). Some of the options and approaches are reviewed and discussed here for characterization, production, and safety assessment of intractable NEPs in GM crops. It is worth noting that these alternative approaches should be applicable to tractable proteins as well, although they can be evaluated using the current protein safety assessment methods.

2.1 Characterization, equivalence, and expression analysis for intractable proteins

This section discusses the challenges and considerations for characterization and protein equivalence studies for surrogate test substances applicable to intractable proteins. It also presents options for test materials and methods for characterization, production, and expression analysis.

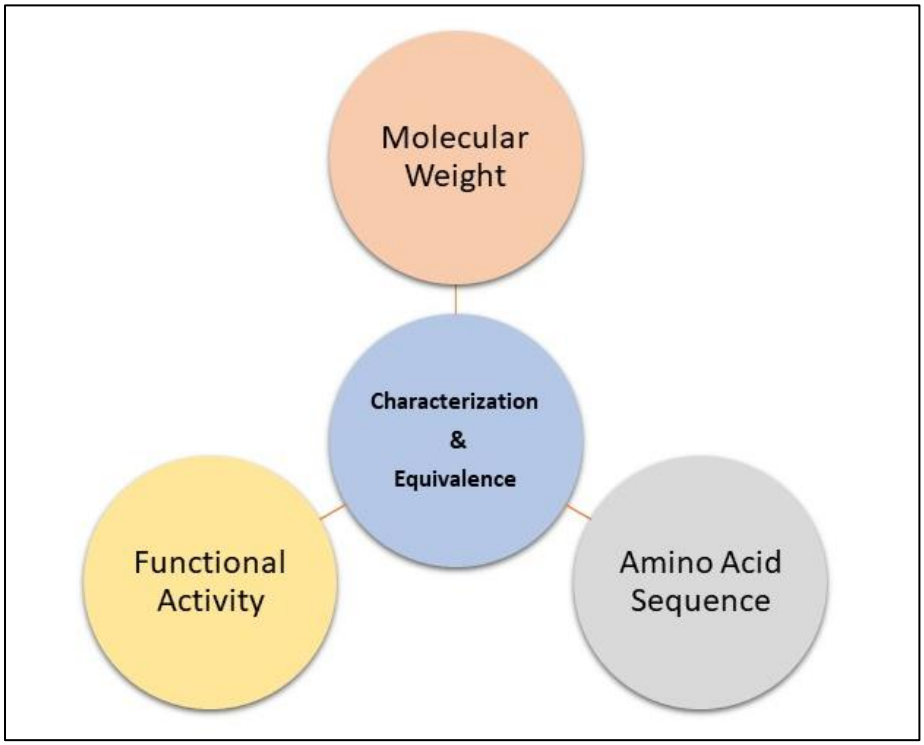

Figure 1. Three core endpoints used to characterize NEPs and assess equivalence of the heterologously produced test protein substance.

2.1 *Protein characterization*

Three core endpoints (i.e., molecular weight, amino acid sequence, and functional activity) have been defined (Brune et al., 2021) for protein characterization and assessment of the physicochemical and functional equivalence of a heterologously produced protein test substance (hereafter, interchangeable with test substance) and the NEP in the GM crop (Figure 1). These characterization endpoints are applicable but need to be considered in the context of the challenges for intractable proteins.

Molecular weight: Molecular weight is a fundamental property of proteins. Techniques such as mass spectrometry, gel electrophoresis (e.g., SDS-PAGE), and western blot can be used to determine the molecular weight of a protein (Edrington et al., 2022; Smith et al., 1990). Not only is the accurate determination of molecular weight for intractable proteins complicated due to factors such as the very low protein yield from plant tissues or heterologous expression systems, but other biochemical properties, such as hydrophobicity of membrane proteins and associated lipids, can affect gel migration in electrophoresis or western blot analyses (Rath et al., 2009). With these challenges in mind, any limited data that can be obtained should be evaluated along with other supporting molecular and biochemical data.

Amino acid sequence: To perform amino acid sequence analysis, a purified or enriched protein preparation is required. Obtaining such preparations can be extremely difficult or impossible due to the low expression levels for intractable proteins such as R-proteins (Habig et al., 2018). Techniques like liquid chromatography-mass spectrometry (LC-MS) and Edman degradation sequencing can be employed to analyze and verify the protein sequence when isolating or enriching the protein is feasible. The extent of sequence coverage obtained through LC-MS peptide mapping depends on factors such as the amount of available protein and the size or ionization efficiency of the peptides from digested proteins. Additionally, the presence of additives resulting from extraction and purification processes, can complicate sequence analysis by LC-MS. Edman sequencing might not be possible due to a blocked N-terminus resulting from modifications like acetylation and challenges of obtaining enough quantity or purity for intractable proteins (Linster & Wirtz, 2018). It is unnecessary and not always practical to determine the complete sequence of the protein for identification or verification purposes. Two or more unique peptide sequences identified by LC-MS/MS are sufficient for confident identification of a protein (Carr et al., 2004; Deutsch et al., 2016; Matsuda et al., 2017). When technical challenges prevent the desired protein sequence information from being obtained, alternative approaches should be considered. For example, confirmation of the gene or RNA transcript sequence from the construct inserted into the GM crop can serve as supporting data for protein sequence verification (Habig et al., 2018).

Protein function: The functional activity of a protein represents another important endpoint for characterization. It is necessary to have functionally active proteins in laboratory-base studies for safety assessment. activity can be impacted or reduced during protein extraction and purification because of factors such as buffer components, pH changes, reactive plant secondary metabolites (e.g., phenols), proteolytic degradation, and removal of lipids for membrane proteins or co-factors for enzymes (Raybould et al., 2013; Sęczyk et al., 2019). Isolating functionally active intractable proteins can be extremely challenging or impossible. For instance, to isolate

glyphosate oxidoreductase (GOX), a membrane protein expressed in glyphosate-tolerant canola, a urea-denaturation step was required for solubilization during purification (EFSA GMO Panel on Genetically Modified Organisms (GMO), 2022). Although the activity of the protein expressed in *Escherichia coli* was detectable in the cell lysate, it could not be demonstrated for the purified, urea denatured GOX produced in 100-gram quantities for safety studies in animals. For membrane associated or transmembrane proteins, microsome fractions prepared from plant tissues or membrane fractions from heterologous expression host cells can be used for *in vitro* functional assays with a potential of retained but reduced activity (Madduri et al., 2012). For some proteins, an *in vitro* activity assay may not be available. In cases where *in vitro* functional activity cannot be directly demonstrated, it may be indirectly supported through the GM crop phenotypes. It had been demonstrated that the functions of disease resistance R-proteins expressed in potatoes were observed through phenotype in the presence of pathogens (Habig et al., 2018).

2.2. Equivalence or sufficient similarity between the protein test substance and the NEP

NEPs in GM crops are often expressed at low levels; therefore, a heterologous (e.g., microbial) system is commonly used to produce a protein test substance for safety assessment studies. Establishing a level of equivalence between the test substance and the NEP in the GM crop is essential. Technical challenges for intractable proteins will likely prevent the equivalence evaluation from following the established paradigm (Bushey et al., 2014; EFSA GMO Panel on Genetically Modified Organisms (GMO), 2020). Equivalence does not imply the NEP and test protein substance are identical, only that the test substance has sufficient biochemical and functional similarity to adequately represent the NEP in safety studies (Brune et al., 2021; Raybould et al., 2013). For example, chloroplast transit peptides (CTPs) are often used for targeting some NEPs into the chloroplast for effective expression and efficacy. The CTPs may be processed as predicted, but not always (McCourt & Duggleby, 2006; Schein et al., 2001). It was observed that extra or different amino acids were present in front of the dicamba monooxygenase (DMO) protein sequence due to different CTP sequences being used for expression in different GM crops. However, the additional amino acids from incomplete processing of the CTP did not impact the protein structure, function, or safety (Wang et al., 2016). Using small or cleavable affinity tags for protein expression and production in heterologous systems can also introduce some minor sequence differences. Affinity tags have been commonly used to aid in protein purification and may be necessary for production of an intractable protein. It has been demonstrated that Mpp75Aa1.1 proteins with or without a His-tag do not show any differences in protein structure, function, or the outcome of a safety assessment (Wang et al., 2022). The isolation of NEPs from plant tissues often needs more complex processes due to very low expression levels. The isolation process often results in reduced functional activity. For example, a 2-fold lower activity in Vpb4Da2 protein isolated from plant tissues was observed compared with heterologously produced Vpb4Da2 (Edrington et al., 2022). The difference and variability in activity would be expected, or possibly more pronounced, for intractable proteins.

The essential elements of assessing sufficient similarity between the NEP in GM crops and the test substance include evaluating the molecular weight, amino acid sequence, and functional activity (Brune et al., 2021) (Figure 1). Due to the challenges of

intractable proteins, it is likely that one or more endpoints to evaluate these characteristics cannot be established in the current way. A more holistic approach is necessary using data that can be practically generated in combination with supporting information such as molecular characterization and mode of action (MOA) data.

Table 1. Examples of Intractable Proteins with Extremely Low Expression Levels in GM Crops[a]

Protein Name	Protein Function	Mean Expression Level in Grain (ppm)	GM Event and Crop	Phenotype	Developer	Year Deregulated by USDA
ZMM28	Transcription factor	0.012	DP202216 Maize	Yield	Pioneer	2020
VNT1	R-protein	<0.5	SPS-000Z6-5 Potato	Late blight protection	Simplot	2020
HaHB4	Transcription factor	0.005	IND-00410-5 Soybean	Yield/abiotic stress tolerance	Verdeca	2019
ATHB17Δ113	Transcription factor	<LOD[b] (LOD=0.00028)	MON 87403 Maize	Yield	Monsanto	2015
BBX32	Transcription factor	<LOD[b] (LOD=0.00025)	MON 87712 Soybean	Yield	Monsanto	2013
PjΔ6D and NcΔ15D	Transmembrane enzyme	1.8 and 10, respectively	MON 87769 Soybean	Improved oil profile	Monsanto	2012
CSPB	Transcription factor	0.072	MON 87460 Maize	Drought tolerance	Monsanto	2011

[a]Data are gathered from the USDA Web site. https://www.aphis.usda.gov/biotechnology/legacy-petition-process/petitions (accessed in March 2024).
[b]LOD refers to the limit of detection.

2.3. Analysis of intractable protein expression levels in crop tissues

The concept of risk can be simply captured in the equation: "Risk = Hazard x Exposure". If a specific hazard is identified, it becomes crucial to determine the impact of the hazard using exposure assessment. Quantifying the level of an intractable protein in a complex matrix poses technical challenges due to solubility issues, background interference with the analytical method, and expression levels that may be below the limit of detection (LOD) (Table 1). Options include targeted protein-specific assays such as ELISA, western blot, or mass spectrometry (MS), depending on factors such as the availability of antibodies, protein standards, or peptide standards. However, using ELISA as a quantitative tool for intractable proteins can be challenging when producing protein standards or high-quality specific antibodies is difficult or impossible. These limitations also apply to protein quantification via western blot. MS can be a suitable choice for quantifying the target protein in a complex matrix, provided that the protein concentration is above the LOD, and matrix interferences are acceptable. A MS method could be applicable for quantification when specific peptides of the NEP from proteolytic digestion are available, separated with liquid chromatography, and ionized for identification and quantification using peptide standards synthesized with stable

isotope labeling (Allen et al., 2014; Tian et al., 2023).

3. *Protein Isolation from plant sources and production with heterologous systems*

This section discusses methods used in protein isolation from plant sources and production with heterologous systems. It also considers the options for protein materials to be used for downstream analyses.

3.1. Isolation of intractable protein from GM plant tissues

Despite the challenges associated with intractable proteins, it is worth assessing the feasibility of isolating a small amount of the NEP from GM plant tissues for characterization. The first step is to consider a protein isolation workflow by evaluating the expression and stability of the intractable NEP in the source material. Depending on the tissue specificity of the promoter used for the NEP expression, certain tissues (e.g., leaf) with higher NEP expression level may be considered and tested for protein isolation. For example, a higher level (0.014 ppm) of ATHB17Δ113 protein was detected in the early growth stage of leaf tissues compared with <0.00028 ppm in grain (Table 1).

Isolation of plant-produced protein often involves multiple extensive separation techniques to enrich a trace amount (e.g., microgram levels) of the NEP, from a large amount of total plant proteins. For example, the average total protein is as much as 365 g/kg soybean seed and 94 g/kg yellow or white corn grain per the USDA database (http://fdc.nal.usda.gov). In addition, other tissue matrices and endogenous compounds such as polyphenols and proteases can interfere with protein purification. For example, gossypol, a major polyphenolic compound in cotton plants, can crosslink free lysine Ɛ-amino groups and cause a decrease in protein yield during purification (Lyman et al., 1959). The typical process for protein isolation from plant sources involves several of the following steps – extraction, clarification, selective precipitation, initial capture, intermediate purification (e.g., antibody or ligand-based affinity chromatography), and final polishing steps such as size exclusion, ion exchange chromatography, or diafiltration (Łącki & Riske, 2020). Antibody-based affinity purification is often a critical step for plant-produced NEP isolation because this is a specific and effective way to isolate a very small quantity (e.g., micrograms) of the protein from a large amount (e.g., kilograms) of plant tissues. However, obtaining sufficient quantity of high-purity NEP in the properly folded form required for antibody generation is challenging because of the characteristics of intractable proteins. Although synthetic peptides could be used as antigens, antibodies generated against peptides tend to have lower specificity and affinity (Maus et al., 2022). These steps often fail for intractable proteins due to their low solubility, instability, or low expression levels (Table 1). Additionally, the long multi-step purification process can impact the protein stability and functional activity (Edrington et al., 2022; Hauser et al., 2008). In general, the more steps in the purification process, the greater the loss of the NEP in quantity and functional activity. The example of isolation of a small amount of cold shock protein B (CSPB) demonstrated the tremendous effort and challenges. The multi-step labor-intensive process includes ammonium sulfate precipitation, diafiltration, anion exchange, antibody affinity chromatography, size exclusion, and final buffer exchange (Wang et al., 2015). The isolation of 60 µg of CSPB protein required 10 kg of ground maize grain, given

its extremely low expression at a level of 0.072 ppm (Table 1).

Although obtaining soluble and active intractable NEPs is very difficult, if not impossible, enriched fractions or partially purified protein samples may be considered for some characterization purposes using techniques such as mass spectrometry, western blot, or *in vitro* activity assays. Enriched fractions can include membrane, chloroplast, or mitochondrion preparations, depending on the protein's subcellular localization. Integral membrane proteins or membrane associated proteins can be enriched by isolating membrane fractions using ultra-speed centrifugation to pellet the membranes and remove most of the soluble proteins from the cellular extracts. The enriched membrane fraction may be used directly or further solubilized in detergent and purified (Bushey et al., 2014). Regardless of the effort put into the process, it is very challenging to isolate enough active membrane proteins to obtain the characterization data described in Section 2.1. If isolating the intractable protein for characterization, such as sequence verification, is impossible owing to extremely low expression (Table 1), protein sequences deduced from cDNA or genomic DNA sequences may be considered and used, along with other available data or information, for safety assessment purposes (Anderson et al., 2019; Habig et al., 2018).

3.2. Production of intractable proteins from heterologous expression systems

Heterologous expression systems are typically used for producing large amounts of protein test substances to take advantage of their capability for over-expression. This is necessary and useful for laboratory studies based on the current regulatory requirements because it is not realistic to isolate the amount of the NEPs from GM crop tissues due to low expression levels. However, the heterologous expression systems may not overcome the challenges associated with intractable proteins, making it very difficult or impossible to generate sufficient amounts of active proteins. Here we discuss some of the challenges and considerations when using heterologous expression and purification methods for intractable proteins.

Heterologous protein expression: The typical workflow for expressing test proteins with heterologous systems involves a multi-step process, including construct design, host selection, expression condition screening, and development of scale-up processes such as fermentation. Construct design includes promoter selection and codon optimization for the host cells for efficient transcription and translation. Affinity fusion tags such as poly-Histidine, glutathione S-transferase, and maltose binding protein (MBP) tags, can also be employed to facilitate purification and improve yield. For example, an N-terminal Histidine tag was used to produce the ZMM28 protein (Anderson et al., 2019). MBP, as a fusion partner, has been reported to exhibit chaperone-like qualities and can improve protein solubility, stability, and proper folding (Kapust & Waugh, 1999).

Heterologous protein expression can be conducted in various host systems, including bacteria (such as *E. coli*, *Bacillus thuringiensis*, and *Pseudomonas fluorescens*), yeast (*Saccharomyces cerevisiae* and *Pichia pastoris*), insect cells (baculovirus-induced expression), mammalian cells, and cell-free expression systems. Each system has its advantages and disadvantages. Microbial expression systems can achieve high expression levels for many proteins. They are also easily scalable, offer many expression vectors, and are relatively cost effective (Bretthauer & Castellino, 1999; Rosano & Ceccarelli, 2014; Rosano et al., 2019). However, bacterial systems might

have challenges when expressing intractable proteins.

Plant-based expression systems, such as transient expression with *Nicotiana benthamiana* or tobacco cell culture, provide plant protein synthesis machinery, protein folding, and post-translational modifications (Yamamoto et al., 2018). However, the transient leaf expression is difficult to scale up. Other eukaryotic expression systems such as yeast, mammalian, and insect cell systems may give better solubility for some proteins but much lower expression level than microbial systems and require larger scales for expression, with the concern that unwanted glycosylation may occur (Kost et al., 2005). Cell-free protein expression systems offer an alternative approach (Garenne et al., 2021). However, some proteins that require chaperones, or that are unstable, may not express well in cell-free systems. The cell-free system also has scalability limitation and is unlikely to produce the gram amounts of proteins required for safety assessments under the current global regulatory paradigm.

Preparation of test proteins expressed with heterologous systems: Obtaining reasonable purity and yield of intractable proteins would be very challenging even with a large amount of cell biomass as starting material. When the expression level of the test protein is low, the relative excess of host proteins makes purification a challenging task. In cases where a suitable amount of intact, soluble, and active intractable protein cannot be produced to conduct laboratory-based studies using traditional methods, alternative approaches can be considered. For example, if a protein is expressed in inclusion bodies or as aggregates, a refolding approach can be tested. Successful refolding procedures require screening of different solubilization and refolding methods, along with optimizing multiple parameters, such as denaturants, pH, salt concentration, divalent ions, polyols, amino acids (e.g., arginine), detergents, sugars, and redox components (Tsumoto et al., 2003). However, these approaches, which are highly depended on protein properties, may, if successful, only produce a very small amount (e.g., nanogram amount) of active proteins.

Transmembrane proteins, known for their low expression levels, poor solubility, and instability due to the hydrophobic nature of membrane-spanning domains, can be challenging to overexpress. Microbial overexpression of a membrane protein is limited due to the surface areas of the host membranes. In addition, the hydrophobicity of membrane proteins leads to protein aggregation, and this can have cytotoxic effects on heterologous host cells, and proteolysis during expression or isolation (Schlegel et al., 2012). Detergents must be used to substitute for the membrane environment for solubilization of the test protein during the extraction and purification process. When the detergent amount is reduced to allow for subsequent analysis and applications, membrane proteins often precipitate. Without the natural membrane environment, active transmembrane proteins are almost impossible to obtain. An enriched membrane fraction preparation which may retain the activity of the membrane protein should be considered for use in the safety assessment. This preparation process has low throughput and low protein yield and purity. The enriched membrane preparation may be used for some characterization and safety assessment studies. For example, enriched microsomes were used for *in vitro* functional assays to demonstrate the similarity of membrane proteins from heterologous expression systems and plant sources (Madduri et al., 2012; Schafer et al., 2016).

In summary, for some intractable proteins, an enriched protein sample may be prepared and

used for certain characterization and laboratory-based studies (Colgrave et al., 2019; Madduri et al., 2012; Schafer et al., 2016). However, producing an adequate quantity of intractable proteins in a purified and active form for traditional safety assessment is essentially impossible. Ultimately, protein-specific factors will determine which, if any, options are feasible.

4. Safety assessment of intractable proteins

A stepwise weight-of-evidence (WOE) approach was established over two decades ago (Codex Alimentarius Commision, 2003, 2009) and has been used to guide the safety assessment for NEPs in GM crops (Anderson et al., 2021; Brune et al., 2021; Bushey et al., 2014; Delaney et al., 2008; Hammond et al., 2013; Roper et al., 2021). Although the current safety evaluation framework has served the purpose thus far, it has become clear that the safety assessment needs to be adapted with a modernized approach based on current knowledge from scientific advancements and from the large body of research and safety data accumulated for NEPs in GM crops (Brune et al., 2021; Waters et al., 2021)

Table 2. Core vs supplemental studies for safety assessment of NEPs in GM crops

Core Studies	Supplemental Studies
• History of safe use (HOSU) of the NEP • HOSU of the source organism • Mode of action (MOA) and functional specificity • Sequence and bioinformatic analysis for toxicity and allergenicity potential • NEP characterization and expression level	• Acute oral toxicity • Protein stability to digestion and heat treatment

A modernized safety evaluation has been proposed for NEPs and includes core assessments for potential hazard identification and supplemental studies that are applied conditionally only when hazards are identified from core studies (Anderson et al., 2021; Brune et al., 2021; McClain et al., 2021; Roper et al., 2021; Waters et al., 2021). Core studies include history of safe use of the NEP, source organism of the gene, MOA and functional specificity of the NEP, sequence analysis against toxins and allergens, and expression level in plants. Supplemental studies may include acute oral toxicity or an exposure assessment (i.e., protein stability to digestion and heat treatment) (Table 2). In the case of environmental risk assessment (ERA) problem formulation process, additional supplementary studies should be considered if a conclusion cannot be made about the pathway to harm using the core study data. One example is the NTO supplementary studies, that may take into consideration of factors such as the NEP expression level, stability, exposure, and spectrum of activity. The stepwise approach with core and supplemental studies is equally appropriate for intractable proteins introduced into GM

crops. However, in some cases, alternative paths or methods may be needed, given the challenges associated with intractable proteins. Below are further discussions on the core studies (section 4.1) and supplemental studies (section 4.2) for intractable proteins.

4.1. Core Studies for potential hazard identification of intractable proteins

4.1.1. History of safe use of the NEP and the source organism

The vast majority of proteins are safe for consumption as macronutrients (Constable et al., 2007). However, a few well-known proteins can be orally toxic, act as antinutrients, or trigger allergies in humans. Examples of these proteins include botulinum neurotoxin (Miyashita et al., 2016), some lectins (Dang & Van Damme, 2015), and the peanut allergen Ara h 2 (Hauser et al., 2008).

During the selection of candidate proteins for use in GM crops, the evaluation of HOSU plays a crucial role (Delaney et al., 2008; Prado et al., 2014). The process begins by investigating the source organism of the gene encoding the protein. The gene can originate from a bacterium, another plant species, or a food crop. Various analyses are conducted to assess the HOSU for food, feed, and environmental safety, and to determine whether the organism or closely related family members contain naturally occurring toxins, anti-nutrients, or allergens. If the source is a microorganism, its pathogenicity or relationship to known pathogens is also evaluated. A candidate gene or protein would likely be dropped from further consideration in early discovery stage if the source organism is clearly hazardous. However, not every protein poses a hazard from a pathogenic bacterial host. Therefore, all available information and knowledge should be considered when assessing the safety of a specific protein, as discussed below. *Bacillus thuringiensis* (*Bt*), a soil-dwelling bacterium, has served as the source of many insect-control proteins (e.g., Cry proteins) that have been introduced into numerous commercially available GM crops. *Bt* has been considered safe due to its widespread presence in the environment, lack of toxicity towards mammals, and long-term use as a natural insect control agent in organic agriculture (Hammond, 2003; Koch et al., 2015).

The HOSU evaluation also determines whether the protein itself or closely related homologs have been safely consumed as human food or animal feed. When the gene encoding the protein is derived from a source organism with some uncertainties or little information available for its HOSU, it does not necessarily mean the protein is a hazard. For example, only about ten proteins out of thousands of proteins expressed in soybeans have been characterized as allergens, indicating a very low likelihood that a NEP originating from soybean is an allergen (Selb et al., 2017; Varunjikar et al., 2023). Therefore, each protein should be evaluated based on its own characteristics in conjunction with other analyses such as MOA, sequence homology and structure familiarity (Capalbo et al., 2019; McClain et al., 2021; Roper et al., 2021).

In case a novel intractable protein does not have a clear HOSU, *per se*, a HOSU may be established if homologs are found in foods that are consumed safely. For example, Habig et al. (2018) reported the safety assessment of the R-protein VNT1. The gene *Rpi-vnt1* encoding the protein was derived from wild potato and introduced into cultivated potato varieties through GM technology to protect against late blight. Other R-protein homologs of VNT1 were found in common food crops including commercial potato and tomato varieties. VNT1 has up to 98% sequence identity with the related R-protein homologs. All these

proteins are consumed safely in cultivated potato, tomato, and pepper, therefore providing support that VNT1 is safe for consumption (Habig et al., 2018).

Another example is the evaluation of the HOSU of maize transcription factor ZMM28 (Anderson et al., 2019). GM maize was developed with increased and extended expression of the endogenous ZMM28 maize protein to enhance grain yield potential. The amino acid sequence of the introduced ZMM28 protein is identical to the native protein in the non-modified maize. Additional evidence indicated that the protein is also found in the grain of several sweet corn varieties and closely related proteins are found in other commonly consumed food crops such as sorghum and rice.

4.1.2. Mode of action and specificity of intractable proteins

The MOA and specificity of a protein contribute to the safety assessment. Knowledge of these characteristics is very important for the WOE-based safety assessment of intractable proteins because of the challenges discussed earlier. The MOA assessment is to investigate how the protein functions and this information can provide some insight into its safety. In addition, information can be generated to show the protein has a defined specificity (e.g., enzymes) or spectrum of activity (e.g., pesticidal proteins). This type of work could be very challenging if purified intractable proteins are needed since isolation of sufficient quantities of the active protein may not be feasible. As an alternative, the MOA might be addressed with information from literature, homologs from another crop or organism, or with information available for another protein within the family with sequence and/or structural similarity (Brune et al., 2021; Capalbo et al., 2019; Moar et al., 2017). To gain additional insight into the MOA or specificity of an intractable protein, other considerations and options may include structural modeling and molecular techniques. The spectrum of activity for a pesticidal protein might be evaluated with the phenotype of the crop (e.g., which pest species are controlled or not controlled). The incorporation of crop tissues or extracts into the media used to conduct laboratory assays against the target or NTOs can also be considered and used.

As an example, VNT1 is a R-protein that recognizes specific effector proteins secreted by plant pathogens and triggers the hypersensitive response of the host plant to limit the pathogen spread (Habig et al., 2018). R-proteins are not pesticidal and do not directly affect pathogens. The proteins specifically impact the host plants through a form of localized programmed cell death at the infection site to protect the plants from further infection or damage. Thus, this MOA has little relevance to toxicity in humans or NTOs. The MOA of the transcription factor ZMM28 was investigated using a combination of *in silico*, laboratory, and field studies (Wu et al., 2019). Based on physiological, biochemical, and molecular characterization, the increased and extended expression of the ZMM28 protein in GM maize enhances photosynthesis and leaf source capacity, which results in plants with enhanced grain yield potential. The MOA and functional specificity of VNT1 and ZMM28 have thus been shown to have no indication of risk to human, animal, or NTO safety.

A clear MOA is an especially important piece of evidence when intractable proteins cannot be isolated in the active form or in sufficient amounts. Phenotypic evidence can be an option to support the function and specificity of the NEPs.

4.1.3. Sequence similarity and bioinformatic analysis for potential toxins and allergens

Amino acid sequence similarity and bioinformatic analysis are important components of the WOE safety assessment for NEPs (Brune et al., 2021; Delaney et al., 2008) including intractable proteins. These analyses are usually done at the early stage of product development to identify any potential risk of toxicity or allergenicity. The protein sequence is compared with known or putative protein toxin sequences for potential similarity using *in silico* bioinformatic tools. In case sequence alignments are observed, further evaluation may be needed to understand more about the domain architecture, structure, and function of the protein, which can help put any sequence similarity into context (Roper et al., 2021). Similarly, the sequence of a protein is compared with the sequences of known or putative allergens in the databases. Examples of publicly available allergen databases are AllergenOnline (http://www.allergenonline.org/) and COMPARE (http://comparedatabase.org) (van Ree et al., 2021). Identification of sequence similarity and structural relatedness with allergens can help determine any needs for supplementary studies to evaluate the relevance of the homology and potential risk (McClain et al., 2021).

4.2. Supplementary studies for characterization of identified hazards

4.2.1. Toxicological assessment for food and feed

The toxicity study for NEPs has traditionally followed the risk assessment protocol for chemicals, using limit doses, despite the significant differences between proteins and chemical compounds. The scientific rational is that if a potential hazard for toxicity is indicated for the intractable protein, additional characterization of the hazard is necessary. Conversely, if core studies reveal no potential toxicity , a routine acute oral toxicity study is not scientifically justified (Roper et al., 2021). Similarly, a 28-day repeated-dose toxicity study is scientifically unjustified when the protein has a HOSU and shows no meaningful homology to known protein toxins (Brune et al., 2021). Proteins that are toxic to mammals are typically cytotoxic, act acutely at high doses, and/or cause damage to epithelial membranes (Roper et al., 2021). Therefore, a 28-day repeated-dose toxicity study does not enhance the safety assessment of proteins beyond what is provided by an acute oral toxicity study (Brune et al., 2021). As previously mentioned, we are not aware of any instance in which a protein has produced adverse effects in the 28-day repeated-dose toxicity study without prior identification of a hazard from *in silico* and *in vitro* data.

As described in previous sections, production of multi-gram quantities of intractable proteins is very challenging if not impossible. Alternative approaches must be considered when a supplementary toxicity study is deemed necessary based on the core assessment for hazard identification. For example, a reduced amount of the protein might be used for the acute oral toxicity study based on a reasonable margin of exposure (e.g., 100-fold the theoretical human exposure) (Hammond et al., 2013; Roper et al., 2021) instead of using the limit dose of 2000 or 5000 mg/kg (protein/body-weight) described in the respective U.S. EPA, OECD and China guidelines (EPA, 2002; Liang et al., 2022; OECD, 2002). The limit dose can be over a thousand to million-fold the exposure to an intractable protein. In many cases, exposure to intractable proteins is negligible due to the very low expression level. It is scientifically justified to dose a protein based on multiples of expected

exposure in toxicity studies rather than an arbitrarily fixed high limit-dose.

If a purified intractable protein cannot be produced in a suitable form for *in-vivo* toxicology studies, other options need to be considered. For example, use of an enriched protein sample or partially purified protein test substance may be necessary to preserve the functional activity of the protein (Bushey et al., 2014). When necessary, diets containing the whole food (e.g., grain) might be considered (Bartholomaeus et al., 2013; Bushey et al., 2014; Hammond et al., 2013; Herman & Ekmay, 2014; Kuiper et al., 2013). Moreover, to reduce animal use, options including estimation of the threshold of toxicological concern (Bushey et al., 2014; Hammond et al., 2013), *in silico* and *in vitro* approaches should be considered.

4.2.2. Environmental risk assessment

The environmental risk assessment (ERA) considers the potential ecological impact of the GM crops or traits on the environment (CFIA, 2017; EPA, 1998). The current framework was developed and adapted from the one used for chemical compounds (EPA, 2002) and may not be entirely suitable when evaluating GM crop traits expressing NEPs. It is critical to use appropriate problem formulation, hypothesis driven and tiered approaches for an ERA (Anderson et al., 2021; Prado et al., 2014). The knowledge of the MOA of the introduced intractable protein can support the hypothesis driven assessment of any potential hazard. If the intractable protein is a plant-incorporated protectant (PIP), such as an insecticidal protein, and a plausible pathway to harm is identified, it is scientifically justified to assess its potential risk to NTOs in the environment. When there is no rational hypothesis to cause harm to NTOs, no studies on NTOs are warranted since the risk is negligible, as is the case for the R-protein VNT1 and transcription factor ZMM28 described above.

Purified NEPs are used for some traditional laboratory NTO studies. For intractable proteins with a potential risk, alternative test substances or study approaches need to be adopted. Instead of using purified proteins, partially purified, enriched protein fractions, or plant tissues should be considered assuming the NTO can tolerate the test material in the diet. It is important to include proper controls in the studies to avoid potential interference from other components in the test substance. If the protein cannot be isolated or enriched from plants, feeding insects directly on plants should be considered as an alternative approach for evaluating potential adverse effects (Romeis et al., 2019).

4.2.3. Stability to digestion and heat treatment

As a part of the current safety assessment for NEPs, *in vitro* digestion studies of the protein by pepsin (simulated gastric fluid) or pancreatin (simulated intestinal fluid) are conducted. The digested samples are analyzed using protein detection methods such as SDS-PAGE and western blot. Digestion with pepsin or pancreatin is typically evaluated with a purified protein. Plant tissue extracts or membrane fractions should be considered as alternative test materials as described previously (Schafer et al., 2016) due to the technical challenges in obtaining purified and functionally active intractable proteins (Bushey et al., 2014; Madduri et al., 2012). Studies directly comparing some purified proteins and the proteins in plant tissue extracts showed that the proteins in purified form and in plant extracts have comparable digestibility (Astwood et al., 1996; Schafer et al., 2016). These observations suggest that digestion studies with plant tissue extracts can be a

suitable alternative, especially for intractable proteins.

Protein stability to digestion has been used as a part of the WOE for the current allergenicity assessment framework by regulatory agencies based on some early studies (Astwood et al., 1996). However, no correlation between protein resistance to gastrointestinal enzymes and allergenic potential has been established (Bogh & Madsen, 2016; Herman et al., 2007; McClain et al., 2021; Wang et al., 2017; Wang et al., 2020). Therefore, digestibility of NEPs does not provide useful information in predicting whether a protein has allergenic potential (Herman et al., 2022; Wang et al., 2021; Wang et al., 2020).

Heat stability testing of NEPs evaluates the impact during food or feed processing by assessing the change of functional activity of the protein through denaturation under increased temperature. It has been previously determined that functional activity and immunoreactivity-based heat stability data does not address protein allergenicity (Privalle et al., 2011). However, the functional activity-based heat stability assessment is relevant in addressing protein structural changes as part of the WOE for toxicity if a potential hazard is identified.

For an intractable protein, enriched protein extracts can be considered and used as alternative testing materials for both digestibility and heat stability. These types of heat lability and digestibility studies may not be possible with intractable proteins in some cases. When a potential hazard is identified using sequence and bioinformatic analysis in the core studies, the digestibility and heat stability data may provide supplemental information relevant to the reduced real-world exposure.

5. Conclusion

Intractable proteins introduced into GM crops present technical challenges in characterization and production of multi-gram quantities to support laboratory-based safety studies under the current regulatory paradigm. The nature of intractable proteins makes it difficult to replicate the studies for current data requirements with the same materials and methods used for tractable proteins. The need for studies should be based on a WOE approach and study designs need to be adaptable on a case-by-case basis. The foundation for characterizing the protein and determining sufficient biochemical and functional similarity between a test substance and the NEP in the GM crop to support safety assessments lies in three core endpoints: molecular weight, amino acid sequence, and functional activity (Brune et al., 2021). The outcome of the core studies including history of safe use, mode of action, functional specificity, sequence, and bioinformatic analysis of the protein, should determine the need for supplemental studies. It is critical to follow this science-based approach to adapt studies for intractable proteins in response to the practical challenges these proteins present. For intractable proteins, if the supplemental studies are deemed necessary, alternative forms of test substances, methods, and doses based on margins of exposure may be used, while still maintaining the rigor of a science-based safety assessment. The considerations and options presented in this review are based on decades of experience and understanding for safety assessment of NEPs in GM crops. With the increasing challenges in developing new GM crop traits, following an updated science-based regulatory safety assessment paradigm is essential to support innovation and delivery of new products, sustain global food security, and reduce agriculture's environmental footprint.

Acknowledgements

The authors thank the Analytical Excellence through Industry Collaboration (AEIC) for the support to prepare and publish this review. The authors also thank Jennifer Anderson, Chad Boeckman, Phil Brune, Kristina Burgin, Luis Burzio, Gerson Graser, Ian Lamb, Eric Ma, Jordan Sottosanto, John Vicini, and Kerrm Yau for their critical review as well as many others who also reviewed the manuscript and provided suggestions. Their contributions are gratefully acknowledged.

References

Allen, D. K., Evans, B. S., & Libourel, I. G. (2014). Analysis of isotopic labeling in peptide fragments by tandem mass spectrometry. PLoS One, 9(3), e91537. https://doi.org/10.1371/journal.pone.0091537

Anderson, J., Bachman, P., Burns, A., Chakravarthy, S., Goodwin, L., Privalle, L., Song, S., & Storer, N. (2021). Streamlining data requirements for the environmental risk assessment of genetically modified (GM) crops for cultivation approvals. Journal of Regulatory Science, 9(1), 26-37. https://doi.org/https://doi.org/10.21423/jrs-v09i1anderson

Anderson, J. A., Brustkern, S., Cong, B., Deege, L., Delaney, B., Hong, B., Lawit, S., Mathesius, C., Schmidt, J., Wu, J., Zhang, J., & Zimmermann, C. (2019). Evaluation of the History of Safe Use of the Maize ZMM28 Protein. Journal of Agricultural and Food Chemistry, 67(26), 7466-7474. https://doi.org/10.1021/acs.jafc.9b00391

Astwood, J. D., Leach, J. N., & Fuchs, R. L. (1996). Stability of food allergens to digestion in vitro. Nature Biotechnology, 14(10), 1269-1273. https://doi.org/10.1038/nbt1096-1269

Bartholomaeus, A., Parrott, W., Bondy, G., & Walker, K. (2013). The use of whole food animal studies in the safety assessment of genetically modified crops: limitations and recommendations. Critical Reviews in Toxicology, 43 Suppl 2(Suppl 2), 1-24. https://doi.org/10.3109/10408444.2013.842955

Bogh, K. L., & Madsen, C. B. (2016). Food Allergens: Is There a Correlation between Stability to Digestion and Allergenicity? Critical Reviews in Food Science and Nutrition, 56(9), 1545-1567. https://doi.org/10.1080/10408398.2013.779569

Bretthauer, R. K., & Castellino, F. J. (1999). Glycosylation of Pichia pastoris -derived proteins. Biotechnology and Applied Biochemistry, 30(3), 193-200. https://doi.org/https://doi.org/10.1111/j.1470-8744.1999.tb00770.x

Brune, P., Chakravarthy, S., Graser, G., Mathesius, C., McClain, S., Petrick, J. S., Sauve-Ciencewicki, A., Schafer, B., Silvanovich, A., Brink, K., Burgin, K., Bushey, D., Cheever, M. L., Edrington, T., Fu, H., Habex, V., Herman, R., Islamovic, E., Lipscomb, E. A.,…Goodwin, L. (2021). Core and Supplementary Studies to Assess the Safety of Genetically Modified (GM) Plants Used for Food and Feed. Journal of Regulatory Science, 9(1), 45-60.

Bushey, D. F., Bannon, G. A., Delaney, B. F., Graser, G., Hefford, M., Jiang, X., Lee, T. C., Madduri, K. M., Pariza, M., Privalle, L. S., Ranjan, R., Saab-Rincon, G., Schafer, B. W., Thelen, J. J., Zhang, J. X., & Harper, M. S. (2014). Characteristics and safety assessment of intractable proteins in genetically modified crops. Regulatory Toxicology and Pharmacology, 69, 154-170. https://doi.org/10.1016/j.yrtph.2014.03.003

Capalbo, D. M. F., Macdonald, P., Fernandes, P. M. B., Rubinstein, C., &

Vicién, C. (2019). Familiarity in the Context of Risk Assessment of Transgenic Crops: Focus on Some Countries in the Americas. Frontiers in Bioengineering Biotechnology, 7, 463. https://doi.org/10.3389/fbioe.2019.00463

Carr, S., Aebersold, R., Baldwin, M., Burlingame, A., Clauser, K., & Nesvizhskii, A. (2004). The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. Molecular & Cellular Proteomics, 3(6), 531-533. https://doi.org/10.1074/mcp.T400006-MCP200

CFIA. (2017). Directive 94-08 (Dir 94-08) Assessment Criteria for Determining Environmental Safety of Plants With Novel Traits. Canadian Food Inspection Agency (CFIA) Retrieved from https://inspection.canada.ca/en/plant-varieties/plants-novel-traits/applicants/directive-94-08

Codex Alimentarius Commision. (2003). Principles for the Risk Analysis of Foods derived from modern biotechnology. Joint FAO/WHO Food Standards Programme, Food and Agriculture Organization of the United Nations, Rome.

Codex Alimentarius Commision. (2009). Food derived from mordern biotechnology. Food and Agriculture Organization of the United Nations and World Health Organization, Rome. Retrived from http://www.fao.org/3/a-a1554e.pdf

Colgrave, M. L., Byrne, K., Caine, J., Kowalczyk, L., Vibhakaran Pillai, S., Dong, B., Lovrecz, G., MacIntosh, S., Scoble, J. A., Petrie, J. R., Singh, S., & Zhou, X. R. (2019). Proteomics reveals the in vitro protein digestibility of seven transmembrane enzymes from the docosahexaenoic acid biosynthesis pathway. Food and Chemical Toxicology, 130, 89-98. https://doi.org/10.1016/j.fct.2019.05.015

Constable, A., Jonas, D. A., Cockburn, A., Davi, A., Edwards, G., Hepburn, P. A., Hérouet-Guicheney, C., Knowles, M., Moseley, B., Oberdörfer, R., & Samuels, F. (2007). History of safe use as applied to the safety assessment of novel foods and foods derived from genetically modified organisms. Food and Chemical Toxicology, 45 12, 2513-2525.

Dang, L., & Van Damme, E. J. M. (2015). Toxic proteins in plants. Phytochemistry, 117, 51-64.

Delaney, B., Astwood, J. D., Cunny, H., Conn, R. E., Herouet-Guicheney, C., Macintosh, S., Meyer, L. S., Privalle, L., Gao, Y., Mattsson, J., & Levine, M. (2008). Evaluation of protein safety in the context of agricultural biotechnology. Food and Chemical Toxicology, 46, S71-S97. https://doi.org/10.1016/j.fct.2008.01.045

Deutsch, E. W., Overall, C. M., Van Eyk, J. E., Baker, M. S., Paik, Y. K., Weintraub, S. T., Lane, L., Martens, L., Vandenbrouck, Y., Kusebauch, U., Hancock, W. S., Hermjakob, H., Aebersold, R., Moritz, R. L., & Omenn, G. S. (2016). Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1. Journal of Proteome Research, 15(11), 3961-3970. https://doi.org/10.1021/acs.jproteome.6b00392

Edrington, T., Wang, R., McKinnon, L., Kessenich, C., Hodge-Bell, K., Li, W., Tan, J., Brown, G., Wang, C., Li, B., & Giddings, K. (2022). Food and feed safety of the Bacillus thuringiensis derived protein Vpb4Da2, a novel protein for control of western corn rootworm. PLoS One, 17(8), e0272311. https://doi.org/10.1371/journal.pone.0272311

EFSA GMO Panel on Genetically Modified Organisms (GMO). (2020). Assessment of genetically modified oilseed rape GT73 for renewal authorisation under Regulation (EC) No 1829/2003 (application EFSA-GMO-RX-002). EFSA Journal Retrieved from https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2020.6199

EFSA GMO Panel on Genetically Modified Organisms (GMO). (2022). Assessment of genetically modified oilseed rape GT73 for renewal authorisation under Regulation (EC) No 1829/2003 (application EFSA-GMO-RX-026/1). EFSA Journal Retrieved from https://efsa.onlinelibrary.wiley.com/doi/org/10.2903/j.efsa.2022.7563

EPA. (1998). United States Environmental Protection Agency (EPA) Guidelines for Ecological Risk Assessment. EPA/630/R-95/002F. Washington, DC. Retrieved from https://www.epa.gov/sites/default/files/2014-11/documents/eco_risk_assessment1998.pdf

EPA. (2002). United States Environmental Protection Agency (EPA)Health Effects Test Guidelines OPPTS 870.1100 Acute Oral Toxicity. Retrieved from https://ntp.niehs.nih.gov/sites/default/files/iccvam/suppdocs/feddocs/epa/epa_870r_1100.pdf

European Commission. (2010). A decade of EU-funded GMO research (2001-2010). Retrieved from https://data.europa.eu/doi/10.2777/97784

Garenne, D., Haines, M. C., Romantseva, E. F., Freemont, P., Strychalski, E. A., & Noireaux, V. (2021). Cell-free gene expression. Nature Reviews Methods Primers, 1(1), 49. https://doi.org/10.1038/s43586-021-00046-x

Habig, J. W., Rowland, A., Pence, M. G., & Zhong, C. X. (2018). Food safety evaluation for R-proteins introduced by biotechnology: A case study of VNT1 in late blight protected potatoes. Regulatory Toxicology and Pharmacology, 95, 66-74. https://doi.org/10.1016/j.yrtph.2018.03.008

Hammond, B. (2003). A Review of the Food/Feed Safety and Benefits of *Bacillus thuringiensis* Protein Containing Insect-Protected Crops. In Agricultural Biotechnology (Vol. 866, pp. 103-123). American Chemical Society. https://doi.org/doi:10.1021/bk-2004-0866.ch00710.1021/bk-2004-0866.ch007

Hammond, B., Kough, J., Herouet-Guicheney, C., & Jez, J. M. (2013). Toxicological evaluation of proteins introduced into food crops. Critical Reviews in Toxicology, 43, 25-42. https://doi.org/10.3109/10408444.2013.842956

Hauser, M., Egger, M., Wallner, M., Wopfner, N., Schmidt, G., & Ferreira, F. (2008). Molecular Properties of Plant Food Allergens: A Current Classification into Protein Families. The Open Immunology Journal, 1, 1-12.

Herman, R. A., & Ekmay, R. (2014). Do whole-food animal feeding studies have any value in the safety assessment of GM crops? Regulatory Toxicology and Pharmacology, 68(1), 171-174. https://doi.org/10.1016/j.yrtph.2013.07.003

Herman, R. A., Woolhiser, M. M., Ladics, G. S., Korjagin, V. A., Schafer, B. W., Storer, N. P., Green, S. B., & Kan, L. (2007). Stability of a set of allergens and non-allergens in simulated gastric fluid. International Journal of Food Sciences and Nutrition, 58(2), 125-141.

Herman, R. A., Zhang, J. X. Q., & Roper, J. M. (2022). Slow alignment of GMO allergenicity regulations with science on protein digestibility. GM Crops & Food, 13(1), 126-130.

https://doi.org/10.1080/21645698.2022.2093552

Kapust, R. B., & Waugh, D. S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Science, 8(8), 1668-1674. https://doi.org/10.1110/ps.8.8.1668

Klümper, W., & Qaim, M. (2014). A meta-analysis of the impacts of genetically modified crops. PLoS One, 9(11), e111629. https://doi.org/10.1371/journal.pone.0111629

Koch, M. S., Ward, J. M., Levine, S. L., Baum, J. A., Vicini, J. L., & Hammond, B. G. (2015). The food and environmental safety of Bt crops [Review]. Frontiers in Plant Science, 6, 283. https://doi.org/10.3389/fpls.2015.00283

Kost, T. A., Condreay, J. P., & Jarvis, D. L. (2005). Baculovirus as versatile vectors for protein expression in insect and mammalian cells. Nature Biotechnology, 23(5), 567-575. https://doi.org/10.1038/nbt1095

Kuiper, H. A., Kok, E. J., & Davies, H. V. (2013). New EU legislation for risk assessment of GM food: no scientific justification for mandatory animal feeding trials. Plant Biotechnology Journal, 11(7), 781-784. https://doi.org/10.1111/pbi.12091

Łącki, K. M., & Riske, F. J. (2020). Affinity Chromatography: An Enabling Technology for Large-Scale Bioprocessing. Biotechnology Journal, 15(1), 1800397. https://doi.org/https://doi.org/10.1002/biot.201800397

Liang, J., Yang, X., Jiao, Y., Wang, D., Zhao, Q., Sun, Y., Li, Y., & Wu, K. (2022). The evolution of China's regulation of agricultural biotechnology. aBIOTECH, 3(4), 237-249. https://doi.org/10.1007/s42994-022-00086-1

Linster, E., & Wirtz, M. (2018). N-terminal acetylation: an essential protein modification emerges as an important regulator of stress responses. Journal of Experimental Botany, 69(19), 4555-4568. https://doi.org/10.1093/jxb/ery241

Lyman, C. M., Baliga, B. P., & Slay, M. W. (1959). Reactions of proteins with gossypol. Archives of Biochemistry and Biophysics, 84(2), 486-497. https://doi.org/https://doi.org/10.1016/0003-9861(59)90610-1

MacIntosh, S. C., Shaw, M., Connelly, M., & Yao, Z. J. (2021). Food and Feed Safety of NS-B5ØØ27-4 Omega-3 Canola (Brassica napus): A New Source of Long-Chain Omega-3 Fatty Acids. Frontiers in Nutrition, 8, 716659. https://doi.org/10.3389/fnut.2021.716659

Madduri, K. M., Schafer, B. W., Hasler, J. M., Lin, G., Foster, M. L., Embrey, S. K., Sastry-Dent, L., Song, P., Larrinua, I. M., Gachotte, D. J., & Herman, R. A. (2012). Preliminary safety assessment of a membrane-bound delta 9 desaturase candidate protein for transgenic oilseed crops. Food and Chemical Toxicology, 50(10), 3776-3784. https://doi.org/10.1016/j.fct.2012.07.013

Matsuda, F., Tomita, A., & Shimizu, H. (2017). Prediction of Hopeless Peptides Unlikely to be Selected for Targeted Proteome Analysis. Mass Spectrometry (Tokyo), 6(1), A0056. https://doi.org/10.5702/massspectrometry.A0056

Maus, A., Renuse, S., Kemp, J., Moehnke, K., Mangalaparthi, K., Chavan, S., Madugundu, A., Vanderboom, P., Dasari, S., Kipp, B., Singh, R., Grebe, S., & Pandey, A. (2022). Comparison of anti-peptide and anti-protein antibody-based purification techniques for detection of SARS-CoV-2 by targeted LC-MS/MS. Advances in Sample

Preparation, 2(100018). https://doi.org/https://doi.org/10.1016/j.sampre.2022.100018

Mbabazi, R., Koch, M., Maredia, K., & Guenthner, J. (2021). Crop Biotechnology and Product Stewardship. GM Crops & Food, 12(1), 106-114. https://doi.org/10.1080/21645698.2020.1822133

McClain, S., Herman, R.A., Islamovic, E., Rakesh, R., Silvanovich, A., Song, P., & Goodwin, L. (2021). Allergy Risk Assessment for Newly Expressed Proteins (NEPs) in Genetically Modified (GM) Plants. Journal of Regulatory Science, 9(1), 67-75.

McCourt, J. A., & Duggleby, R. G. (2006). Acetohydroxyacid synthase and its role in the biosynthetic pathway for branched-chain amino acids. Amino Acids, 31(2), 173-210. https://doi.org/10.1007/s00726-005-0297-3

McHughen, A., & Smyth, S. (2008). US regulatory system for genetically modified [genetically modified organism (GMO), rDNA or transgenic] crop cultivars. Plant Biotechnology Journal, 6(1), 2-12. https://doi.org/10.1111/j.1467-7652.2007.00300.x

Miyashita, S.-I., Sagane, Y., Suzuki, T., Matsumoto, T., Niwa, K., & Watanabe, T. (2016). "Non-Toxic" Proteins of the Botulinum Toxin Complex Exert In-vivo Toxicity. Scientific Reports, 6(1), 31043. https://doi.org/10.1038/srep31043

Moar, W. J., Evans, A. J., Kessenich, C. R., Baum, J. A., Bowen, D. J., Edrington, T. C., Haas, J. A., Kouadio, J. K., Roberts, J. K., Silvanovich, A., Yin, Y., Weiner, B. E., Glenn, K. C., & Odegaard, M. L. (2017). The sequence, structural, and functional diversity within a protein family and implications for specificity and safety: The case for ETX_MTX2 insecticidal proteins. Journal of Invertebrate Pathology, 142, 50-59. https://doi.org/10.1016/j.jip.2016.05.007

National Academies of Sciences, Engineering, and Medicine. (2016). Genetically Engineered Crops: Experiences and Prospects. The National Academies Press. https://doi.org/10.17226/23395

Nicolia, A., Manzo, A., Veronesi, F., & Rosellini, D. (2014). An overview of the last 10 years of genetically engineered crop safety research. Critical Reviewers in Biotechnology, 34, 77-88.

OECD. (1993). Safety Evaluation of Foods Derived by Modern Biotechnology: Concepts and Principles. Paris, France: Organisation for Economic Co-operation and Development (OECD)

OECD. (2002). Guidelines for Testing of Chemicals, Section 423, Acute Oral Toxicity – Acute Toxic Class Method. Organisation for Economic Co-operation and Development Retrieved from https://ntp.niehs.nih.gov/sites/default/files/iccvam/suppdocs/feddocs/oecd/oecd_gl423.pdf

Prado, J. R., Segers, G., Voelker, T., Carson, D., Dobert, R., Phillips, J., Cook, K., Cornejo, C., Monken, J., Grapes, L., Reynolds, T., & Martino-Catt, S. (2014). Genetically engineered crops: from idea to product. Annual Review of Plant Biology, 65, 769-790. https://doi.org/10.1146/annurev-arplant-050213-040039

Privalle, L., Bannon, G., Herman, R., Ladics, G., McClain, S., Stagg, N., Ward, J., & Herouet-Guicheney, C. (2011). Heat stability, its measurement, and its lack of utility in the assessment of the potential allergenicity of novel proteins. Regulatory Toxicology and Pharmacology, 61(3), 292-295. https://doi.org/10.1016/j.yrtph.2011.08.009

Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., & Deber, C. M. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proceedings of National Academy of Sciences of the United States of America, 106(6), 1760-1765. https://doi.org/10.1073/pnas.0813167106

Raybould, A., Kilby, P., & Graser, G. (2013). Characterising microbial protein test substances and establishing their equivalence with plant-produced proteins for use in risk assessments of transgenic crops. Transgenic Research, 22, 445-460. https://doi.org/10.1007/s11248-012-9658-3

Romeis, J., Naranjo, S. E., Meissle, M., & Shelton, A. M. (2019). Genetically engineered crops help support conservation biological control. Biological Control, 130, 136-154. https://doi.org/10.1016/j.biocontrol.2018.10.001

Roper, J., Lipscomb, E. A., Petrick, J. S., Ranjan, R., Sauve-Ciencewicki, A., & Goodwin, L. (2021). Toxicological Assessment of Newly Expressed Proteins (NEPs) in Genetically Modified (GM) Plants. Journal of Regulatory Science, 9(1), 61-66. https://doi.org/ https://doi.org/10.21423/jrs-v09i1roper

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. Frontiers in Microbiology, 5, 172. https://doi.org/10.3389/fmicb.2014.00172

Rosano, G. L., Morales, E. S., & Ceccarelli, E. A. (2019). New tools for recombinant protein production in Escherichia coli: A 5-year update. Protein Science, 28(8), 1412-1422. https://doi.org/10.1002/pro.3668

Schafer, B. W., Embrey, S. K., & Herman, R. A. (2016). Rapid simulated gastric fluid digestion of in-seed/grain proteins expressed in genetically engineered crops. Regulatory Toxicology and Pharmacology, 81, 106-112. https://doi.org/10.1016/j.yrtph.2016.08.004

Schein, A. I., Kissinger, J. C., & Ungar, L. H. (2001). Chloroplast transit peptide prediction: a peek inside the black box. Nucleic Acids Research, 29(16), E82. https://doi.org/10.1093/nar/29.16.e82

Schlegel, S., Löfblom, J., Lee, C., Hjelm, A., Klepsch, M., Strous, M., Drew, D., Slotboom, D. J., & de Gier, J. W. (2012). Optimizing membrane protein overexpression in the Escherichia coli strain Lemo21(DE3). Journal of Molecular Biology, 423(4), 648-659. https://doi.org/10.1016/j.jmb.2012.07.019

Sęczyk, Ł., Świeca, M., Kapusta, I., & Gawlik-Dziki, U. (2019). Protein–Phenolic Interactions as a Factor Affecting the Physicochemical Properties of White Bean Proteins. Molecules, 24(3). https://doi.org/10.3390/molecules24030408

Selb, R., Wal, J. M., Moreno, F. J., Lovik, M., Mills, C., Hoffmann-Sommergruber, K., & Fernandez, A. (2017). Assessment of endogenous allergenicity of genetically modified plants exemplified by soybean - Where do we stand? Food and Chemical Toxicology, 101, 139-148. https://doi.org/10.1016/j.fct.2017.01.014

Sharma, P., Singh, S. P., Iqbal, H. M. N., Parra-Saldivar, R., Varjani, S., & Tong, Y. W. (2022). Genetic modifications associated with sustainability aspects for sustainable developments. Bioengineered, 13(4), 9508-9520. https://doi.org/10.1080/21655979.2022.2061146

Skinner, W. S., Phinney, B. S., Herren, A., Goodstal, F. J., Dicely, I., & Facciotti, D. (2016). Using LC-MS Based Methods for Testing the Digestibility of a Nonpurified Transgenic Membrane Protein in Simulated

Gastric Fluid. Journal of Agricultural and Food Chemistry, 64(25), 5251-5259. https://doi.org/10.1021/acs.jafc.6b01829

Smith, R. D., Loo, J. A., Edmonds, C. G., Barinaga, C. J., & Udseth, H. R. (1990). New developments in biochemical mass spectrometry: electrospray ionization. Analytical Chemistry, 62(9), 882-899. https://doi.org/10.1021/ac00208a002

Tian, X., Permentier, H. P., & Bischoff, R. (2023). Chemical isotope labeling for quantitative proteomics. Mass Spectrometry Review, 42(2), 546-576. https://doi.org/10.1002/mas.21709

Tsumoto, K., Ejima, D., Kumagai, I., & Arakawa, T. (2003). Practical considerations in refolding proteins from inclusion bodies. Protein Expression and Purification, 28(1), 1-8. https://doi.org/10.1016/s1046-5928(02)00641-1

van Ree, R., Sapiter Ballerda, D., Berin, M. C., Beuf, L., Chang, A., Gadermaier, G., Guevera, P. A., Hoffmann-Sommergruber, K., Islamovic, E., Koski, L., Kough, J., Ladics, G. S., McClain, S., McKillop, K. A., Mitchell-Ryan, S., Narrod, C. A., Pereira Mouriès, L., Pettit, S., Poulsen, L. K.,…Bowman, C. (2021). The COMPARE Database: A Public Resource for Allergen Identification, Adapted for Continuous Improvement. Frontiers in Allergy, 2, 700533. https://doi.org/10.3389/falgy.2021.700533

Varunjikar, M. S., Bøhn, T., Sanden, M., Belghit, I., Pineda-Pampliega, J., Palmblad, M., Broll, H., Braeuning, A., & Rasinger, J. D. (2023). Proteomics analyses of herbicide-tolerant genetically modified, conventionally, and organically farmed soybean seeds. Food Control, 151, 109795. https://doi.org/10.1016/j.foodcont.2023.109795

Wang, C., Bean, G. J., Chen, C. J., Kessenich, C. R., Peng, J., Visconti, N. R., Milligan, J. S., Moore, R. G., Tan, J., Edrington, T. C., Li, B., Giddings, K. S., Bowen, D., Luo, J., Ciche, T., & Moar, W. J. (2022). Safety assessment of Mpp75Aa1.1, a new ETX_MTX2 protein from Brevibacillus laterosporus that controls western corn rootworm. PLoS One, 17(9), e0274204. https://doi.org/10.1371/journal.pone.0274204

Wang, C., Burzio, L. A., Koch, M. S., Silvanovich, A., & Bell, E. (2015). Purification, characterization and safety assessment of the introduced cold shock protein B in DroughtGard maize. Regulatory Toxicology and Pharmacology, 71(2), 164-173. https://doi.org/10.1016/j.yrtph.2014.12.014

Wang, C., Glenn, K. C., Kessenich, C., Bell, E., Burzio, L. A., Koch, M. S., Li, B., & Silvanovich, A. (2016). Safety assessment of dicamba mono-oxygenases that confer dicamba tolerance to various crops. Regulatory Toxicology and Pharmacology, 81, 171-182. https://doi.org/https://doi.org/10.1016/j.yrtph.2016.08.014

Wang, R., Edrington, T. C., Storrs, S. B., Crowley, K. S., Ward, J. M., Lee, T. C., Liu, Z. L., Li, B., & Glenn, K. C. (2017). Analyzing pepsin degradation assay conditions used for allergenicity assessments to ensure that pepsin susceptible and pepsin resistant dietary proteins are distinguishable. PLoS One, 12(2), e0171926. https://doi.org/10.1371/journal.pone.0171926

Wang, R., Houston, N., Cheever, M., Geng, T., Gillikin, N., McDonald, J., Serrano, H., Shippar, J., Tetteh, A., Wang, Y., & Liu, Z. L. (2021). Can Mass Spectrometry Analysis of In Vitro Digestion Products Improve the Assessment of Allergenic Potential of a

Newly Expressed Protein? Journal of Regulatory Science, 9(1), 76-83.

Wang, R., Wang, Y., Edrington, T. C., Liu, Z., Lee, T. C., Silvanovich, A., Moon, H. S., Liu, Z. L., & Li, B. (2020). Presence of small resistant peptides from new in vitro digestion assays detected by liquid chromatography tandem mass spectrometry: An implication of allergenicity prediction of novel proteins? PLoS One, 15(6), e0233745. https://doi.org/10.1371/journal.pone.0233745

Waters, S., Ramos, A., Culler, A. H., Hunst, P., Zeph, L., Gast, R., Mahadeo, D., Sottosanto, J., Huber, S., Shan, G., Chakravarthy, S., & Goodwin, L. (2021). Recommendations for science-based safety assessment of genetically modified (GM) plants for food and feed uses. Journal of Regulatory Science, 9(1), 16-21.

Wu, J., Lawit, S. J., Weers, B., Sun, J., Mongar, N., Van Hemert, J., Melo, R., Meng, X., Rupe, M., Clapp, J., Haug Collet, K., Trecker, L., Roesler, K., Peddicord, L., Thomas, J., Hunt, J., Zhou, W., Hou, Z., Wimmer, M.,…Habben, J. E. (2019). Overexpression of zmm28 increases maize grain yield in the field. Proceedings of National Academy of Sciences of the United States of America, 116(47), 23850-23858. https://doi.org/10.1073/pnas.1902593116

Yamamoto, T., Hoshikawa, K., Ezura, K., Okazawa, R., Fujita, S., Takaoka, M., Mason, H. S., Ezura, H., & Miura, K. (2018). Improvement of the transient expression system for production of recombinant proteins in plants. Scientific Reports, 8(1), 4755. https://doi.org/10.1038/s41598-018-23024-y