Regulatory Toxicology and Pharmacology 69 (2014) 154–170

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Characteristics and safety assessment of intractable proteins in genetically modified crops

Dean F. Bushey [a,*], Gary A. Bannon [b], Bryan F. Delaney [c], Gerson Graser [d], Mary Hefford [e], Xiaoxu Jiang [f], Thomas C. Lee [b], Krishna M. Madduri [g], Michael Pariza [h], Laura S. Privalle [f,1], Rakesh Ranjan [a], Gloria Saab-Rincon [i], Barry W. Schafer [g], Jay J. Thelen [j], John X.Q. Zhang [c], Marc S. Harper [c,2]

[a] Bayer CropScience, 2 T.W. Alexander Drive, P.O. Box 12014, Research Triangle Park, NC 27709-2014, USA
[b] Monsanto Company, Global Regulatory, 800 North Lindbergh Boulevard, St. Louis, MO 63167, USA
[c] DuPont Pioneer, 7300 NW 62nd Avenue, P.O. Box 1004, Johnston, IA 50131, USA
[d] Syngenta Biotechnology, Inc., P.O. Box 12257, 3054 East Cornwallis Road, Research Triangle Park, NC 27709-2257, USA
[e] Centre for Biologics Research, Sir Frederick G. Banting Research Centre, 251 Promenade Sir Frederick Banting Driveway, Ottawa, Ontario K1A 0K9, Canada
[f] BASF Plant Science, L.P., Regulatory Science, 26 Davis Drive, Research Triangle Park, NC 27709-3528, USA
[g] Dow AgroSciences, 9330 Zionsville Road, Indianapolis, IN 46268, USA
[h] Department of Food Science, University of Wisconsin-Madison, 1550 Linden Drive, Madison, WI 53706, USA
[i] Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, AP 510-3, CP 62250 Cuernavaca, Morelos, Mexico
[j] University of Missouri, Division of Biochemistry, Christopher S. Bond Life Sciences Center, Columbia, MO 65211, USA

ARTICLE INFO

Article history:
Received 7 October 2013
Available online 22 March 2014

Keywords:
Genetically modified (GM) crops
Intractable proteins
Food safety
Membrane proteins
Signaling proteins
Transcription factors
N-glycosylated proteins
Resistance proteins
Weight-of-evidence approach

ABSTRACT

Genetically modified (GM) crops may contain newly expressed proteins that are described as "intractable". Safety assessment of these proteins may require some adaptations to the current assessment procedures. Intractable proteins are defined here as those proteins with properties that make it extremely difficult or impossible with current methods to express in heterologous systems; isolate, purify, or concentrate; quantify (due to low levels); demonstrate biological activity; or prove equivalency with plant proteins. Five classes of intractable proteins are discussed here: (1) membrane proteins, (2) signaling proteins, (3) transcription factors, (4) N-glycosylated proteins, and (5) resistance proteins (R-proteins, plant pathogen recognition proteins that activate innate immune responses). While the basic tiered weight-of-evidence approach for assessing the safety of GM crops proposed by the International Life Sciences Institute (ILSI) in 2008 is applicable to intractable proteins, new or modified methods may be required. For example, the first two steps in Tier I (hazard identification) analysis, gathering of applicable history of safe use (HOSU) information and bioinformatics analysis, do not require protein isolation. The extremely low level of expression of most intractable proteins should be taken into account while assessing safety of the intractable protein in GM crops. If Tier II (hazard characterization) analyses requiring animal feeding are judged to be necessary, alternatives to feeding high doses of pure protein may be needed. These alternatives are discussed here.

Abbreviations: BW, body weight; CAH1, chloroplast-localized carbonic anhydrase; CC, coiled-coil; CMC, critical micelle concentration; CO, CONSTANS; CYP, cytochrome P450; D9DS, delta 9 desaturase; DGAT, diacylglycerol acyltransferase; EPSPS, 5-enol-pyruvylshikimate-3-phosphate synthase; ER, endoplasmic reticulum; ETI, effector-triggered immunity; FARRP, Food Allergy Research and Resource Program; GFP, green fluorescent protein; GI, gastrointestinal; GM, genetically modified; GPI, glucosyl phosphatidyl inositide; HLB, hydrophilic–lipophilic balance; HOSU, history of safe use; IFBiC, International Food Biotechnology Committee; ILSI, International Life Sciences Institute; IV, intravenous; LRR, leucine-rich repeat; NBS, nucleotide-binding site; NOAEL, no-observed-adverse-effect level; OECD, Organisation for Economic Co-operation and Development; PAMP, pathogen-associated molecular pattern; PAT, phosphinothricin acetyltransferase; PEPCK, phosphoenolpyruvate carboxylase kinase; PHA, phytohemagglutinin; PTI, PAMP-triggered immunity; R-proteins, plant pathogen recognition proteins that activate innate immune responses; RLK, receptor-like kinase; RLP, receptor-like protein; SDS, sodium dodecyl sulfate; SGF, simulated gastric fluid; SIF, simulated intestinal fluid; STAND, signal transduction ATPase with numerous domains; TAG, triacylglycerol; TFAPs, transcription factor accessory proteins; TTC, threshold of toxicological concern; TIR, Toll and interleukin-1 receptor.

* Corresponding author. Fax: +1 919 549 2892.
E-mail address: dean.bushey@bayer.com (D.F. Bushey).
[1] Present address: Bayer CropScience, 2 T.W. Alexander Drive, P.O. Box 12014, Research Triangle Park, NC 27709-2014, USA.
[2] Present address: Syngenta Biotechnology, Inc., P.O. Box 12257, 3054 East Cornwallis Road, Research Triangle Park, NC 27709-2257, USA.

http://dx.doi.org/10.1016/j.yrtph.2014.03.003
0273-2300/

1. Introduction

The safety of proteins expressed in genetically modified (GM) crops (hereafter "transgenic proteins") has been assessed ever since scientists first had the ability to introduce genes into crops. There are many opinions on what tests or studies should be required to scientifically document that transgenic proteins are safe, especially when the protein has a history of safe use (HOSU) or when the expression level is very low.

For the transgenic proteins that are expressed in GM crops today, a comparative safety assessment process was implemented in the 1990s in which scientific studies were carried out to identify the similarities and differences between a newly developed GM crop and its conventional (non-GM) counterpart that had a HOSU. This approach assesses (1) the agronomic/morphological characteristics of the GM crop; (2) macro- and micronutrient composition and content of important anti-nutrients and toxicants; (3) molecular characteristics, protein expression, and safety of the newly introduced protein(s) and their metabolites; and, if appropriate, (4) the nutritional characteristics of the novel product compared with that of its conventional counterpart, by testing wholesomeness in animal models (e.g., poultry feeding studies). Any identified biological differences are assessed further to determine whether safety issues or concerns exist and then to evaluate the associated risk. This comparative safety assessment process is also known as substantial equivalence. This approach has become standard for safety evaluation of GM crops and has been described in multiple publications (e.g., Delaney et al., 2008; Hammond, 2008). A more recent work examined whether the same approach was sufficient to demonstrate the safety of GM crops that have improved function by altering endogenous gene expression via RNAi technology or expression of transcription factors (Parrott et al., 2010). However, most of the transgenic proteins currently in GM crops are foreign to the target plant and either toxic to insects or afford tolerance to commercial herbicides. More importantly, they are amenable to production of significant amounts in heterologous systems, isolation, and subsequent testing. Some proteins from the next generation of transgenic crops are already proving to be much more difficult to study. In some cases the transgenic protein will be an integral part of the substructure of the plant cell, in others it may be closely related to a protein of the target plant, and in still others it may be present in the target plant, but be expressed ectopically in the GM crop. Some proteins may only exist at very low levels for a short time and be hard to detect and/or identify in the plant.

There are important questions associated with these "intractable" proteins. Do these proteins pose a safety issue? Do they need to be regulated? If necessary, how does one perform a safety assessment on intractable proteins? While it is unlikely that a protein that is unstable outside of its normal plant environment can be toxic to an animal or human, there will be intractable proteins for which a safety assessment is appropriate, or requested by regulators. In these cases, there are studies that can contribute to the safety assessment of intractable proteins should it be necessary, and those studies are the focus of this paper.

Intractable proteins are defined here as those proteins with properties that make it impossible or extremely difficult to (1) express in a heterologous system; (2) quantify (due to low levels); (3) isolate, concentrate, or purify from either heterologous expression systems or the GM plant; (4) demonstrate functionality of the isolated protein; or (5) prove equivalency of the heterologously produced protein with the plant-expressed protein. These limitations are important, because in 2008, a document jointly published by the International Life Sciences Institute (ILSI) and the International Food Biotechnology Committee (IFBiC) recommended a systematic weight-of-evidence tiered approach to assess the safety of novel proteins expressed in GM crops (Delaney et al., 2008). Safety evaluation of the candidate novel protein begins with a Tier I potential hazard identification, which includes HOSU, bioinformatics analysis, mode of action, *in vitro* digestibility and stability in the presence of simulated gastric fluid (SGF) and simulated intestinal fluid (SIF), expression level, and dietary intake evaluation. If these elements of the safety evaluation are satisfactory, then it can be concluded that the protein is safe to express in GM crops. If, however, the safety of the protein cannot be confirmed in the Tier I analyses, then Tier II hazard characterization studies should be considered. These Tier II studies could include acute and possibly repeated-dose toxicity studies in mice or rats and, if warranted, hypothesis-based evaluations. Some of the Tier I tests and all of the Tier II hazard tests require grams of protein. However, production, isolation, or concentration of sufficient quantities of functionally active proteins for use in safety studies may not be possible. For example, integral membrane proteins are not only difficult to express in heterologous systems, they (due to their hydrophobicity) have very limited solubility in the types and levels of vehicles that would be appropriate for toxicity testing. Many proteins form suspensions at high concentrations, but membrane proteins oligomerize into uncharacterizable forms. The question then becomes, how does one provide appropriate scientific data to support the safety assessment of a GM crop that contains an intractable protein?

Classes of proteins in which all or at least some of the proteins might be intractable include (1) membrane proteins, (2) signaling proteins, (3) transcription factors, (4) N-glycosylated proteins, and (5) resistance (R)-proteins. The characteristics that could render each type of protein intractable are discussed along with tools and science-based solutions for safety assessment of intractable proteins. The scope of this paper focuses only on the safety assessment and functionality of the intractable protein itself and not the safety of the crop containing the protein.

2. Classes of intractable proteins

Table 1 summarizes the classes of intractable proteins discussed in this section.

2.1. Membrane proteins

2.1.1. Definition of membrane proteins

"Membrane protein" is a biochemical term used to describe polypeptides that associate with lipid membranes, either stably

Table 1
Overview of issues associated with different intractable protein classes.

Issue	Protein class				
	Membrane proteins	Signaling proteins	Transcription factors	N-glycosylated proteins	R-proteins
Absence of suitable heterologous expression system	✓		✓	✓	
Low level of expression in GM crop	✓	✓	✓		✓
Inability to test the functionality of isolated protein	✓				✓
Inability to determine equivalence of heterologously produced protein and plant-expressed protein	✓	✓	✓	✓	✓

or transiently. Membrane proteins perform a diverse set of cellular functions ranging from metabolite exchange to cell signaling, and their interaction with the membrane is critical to their biological activity and function. They can act as enzymes, electron carriers, ion channels, transporters, pumps, photosynthetic reaction centers, or receptors. Because of their involvement in many critical cellular functions, select membrane proteins are of major interest for introduction into GM crops.

Classification of membrane proteins is based on the physical properties of the protein, which can be ascertained either by structural prediction using bioinformatics (e.g., hydrophobicity plots) or by experimental approaches.

The mechanisms used by proteins to associate with membranes are diverse and therefore necessitate sub-classification. Based on their location or interaction with the biomembrane, membrane proteins can be integral, peripheral, or lipid-anchored (Fig. 1). Examples of proteins in each class are provided in Table 2.

2.1.2. Integral membrane proteins

An integral membrane protein is a protein that spans (or is embedded into) a lipid bilayer, such as the plasma, nuclear, or organellar envelope membrane. Integral membrane proteins constitute a considerable proportion of all membrane proteins. They are involved in various cellular tasks such as catalysis, membrane transport, and cell signaling. Integral membrane proteins are intrinsically aqueous-insoluble and recalcitrant to solubilization with alkaline carbonate or high-salt conditions due to direct hydrophobic associations with the aliphatic portion of a lipid membrane. Solubilization of integral membrane proteins requires disruption of the membrane, typically through the use of detergents, to create an artificial micellar environment for the protein (Henningsen et al., 2002; Macher and Yen, 2007). The types of surfactant(s) (non-ionic, zwitterionic, or ionic) required for complete solubilization reflect the degree of membrane association (Lin and Guidotti, 2009). Ionic detergents such as sodium dodecyl sulfate (SDS) are the most disruptive detergents, many of which have denaturing properties at high concentrations and temperatures.

2.1.3. Peripheral membrane proteins

Peripheral membrane proteins are associated with cellular membranes *via* direct or indirect association with either integral membrane protein(s) or lipids (Marsh et al., 2002). While some peripheral membrane proteins may transiently interact with membranes, many interactions are considered stable, and these stable interactions can be hydrophobic, ionic, or covalent. This association occurs *in vivo,* and, when the interaction is stable, under standard protein extraction conditions *in vitro*. These proteins can be stripped from membranes by washing with alkaline carbonate or high-ionic-strength solutions/buffers. Cell signaling and other events can be controlled by reversible attachment of proteins to membranes. Membrane binding may also bring together an enzyme and its substrates or cause conformational changes in a protein that lead to its activation.

2.1.4. Lipid-anchored proteins

Perhaps the most difficult membrane proteins to classify solely from the amino acid sequence are the lipid-anchored proteins. These are generally soluble proteins that are modified post-translationally

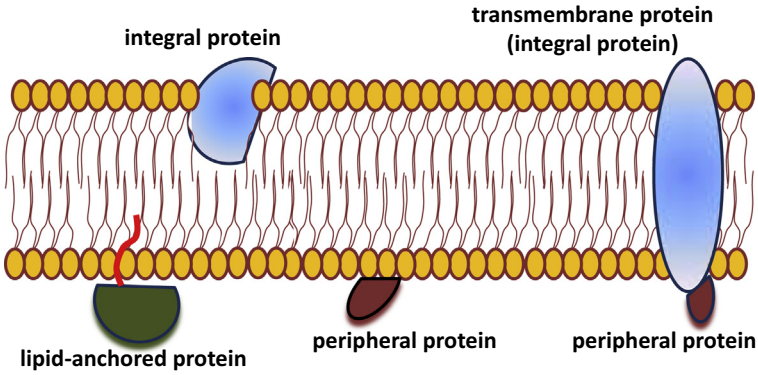

Fig. 1. Schematic figure showing the different classes of membrane proteins.

Table 2
Categories and examples of membrane proteins.

Category	Examples
Integral membrane proteins	
Enzymes	Fatty acid desaturase
Transporters	Glucose transporter
Adhesion proteins	Integrins
Receptor proteins	Nicotinic acetylcholine receptor
Ion channels	Calcium channel
Multi-enzyme complexes	Cytochrome P450 (CYP) enzymes in cooperation with NADPH-cytochrome P450 oxidoreductase (reductase)
Peripheral membrane proteins	
Enzymes	Phospholipases, cholesterol oxidases, glycosyltransferase, palmitoyl protein thioesterases
Structural domains that mediate attachment of other proteins	Annexins
Transporters	Polyisoprenoid-binding protein
Electron carriers	Cytochrome C, cupredoxins, nitrite reductase, some flavoproteins
Polypeptide hormones	Insulin, antimicrobial peptides
Lipid-anchored proteins	
Glycosylphosphatidylinositol (GPI)-anchored proteins	
Prenylated proteins	G-proteins, nuclear lamins
Acylated proteins	Src family of tyrosine kinases

with an aliphatic group that enables the protein to associate with membranes in a semi-reversible manner. The types of post-translational modifications that are capable of making a protein amphipathic include, but are not limited to, acylation (e.g., myristoylation, palmitoylation; Charollais and Van Der Goot, 2009), prenylation and farnesylation (Resh, 1996), and glucosyl phosphatidyl inositide (GPI) anchoring (Paulick and Bertozzi, 2008).

2.1.5. Intractability issues associated with membrane proteins

Membrane proteins are considered intractable because of the difficulty in developing a suitable heterologous environment for expression of the recombinant protein, purification of the protein from the membrane fraction, and/or retention of biological activity once a purified (or enriched) state is achieved. Most membrane proteins tend to be a minor component of the cell mass, making isolation from plant cells difficult and further complicating the production of large quantities of these proteins in heterologous systems. Some integral membrane proteins are part of protein complexes, which pose additional challenges for heterologous and functional expression (Roos et al., 2012), as it is necessary to simulate the micro-environment of this highly organized system. Estimating the activity of the membrane protein when part of a multi-enzyme complex and comparing this activity when the membrane protein is expressed in different hosts is a major challenge.

2.1.5.1. Heterologous over-expression of membrane proteins. Due to the specific nature of membrane proteins, their heterologous expression often results in a very limited, or undetectable amount, of functional protein. If the protein is present at all, it may be produced as an insoluble form in inclusion bodies, and would therefore be nonfunctional. The main factor limiting the over-expression of membrane proteins is the availability of membrane space within the expression host, resulting in lower concentration compared to soluble proteins expressed in the cytoplasm. Even though the presence of membrane proteins is often tested in *Escherichia coli*, eukaryotic hosts such as yeast, mammalian, and insect cells are more likely to be useful for obtaining higher levels of functional membrane protein (see Section 2.4.3). In particular, eukaryotic expression hosts such as yeast (*Pichia pastoris*) and insect (*Spodoptera frugiperda*) are very useful for membrane protein evaluation (André et al., 2006; Asada et al., 2011; Bernaudat et al., 2011). Photosynthetic bacteria such as *Rhodobacter sphaeroides* also provide a similar advantage, as it is possible to create more membrane space for heterologous membrane proteins by systematically disrupting components of the photosynthetic complex, which is membrane bound. This strategy has been successfully used to express numerous proteins (Jaschke et al., 2011; Laible et al., 2004; Roy et al., 2008). Similarly, cell-free protein production systems can be exploited by supplementing the reaction with liposomes.

Unlike soluble proteins, membrane proteins have to be properly inserted into lipid bilayers for them to take on a proper conformation. The presence of multiple membrane-spanning domains further complicates heterologous expression. Although a protein with only a single membrane-anchoring domain can be expressed in a soluble form by deleting the domain, it would no longer be functionally equivalent to the introduced protein in a GM crop. Furthermore, the assembly of membrane proteins and protein complexes requires specific supporting operations, such as membrane insertion and protein folding facilitated by translocon complexes (Bowie, 2005; Skach, 2009). Differences in or absence of the supporting elements in the heterologous host complicate expression.

2.1.5.2. Purification of heterologously expressed membrane proteins. The techniques used to purify soluble proteins, such as immunoaffinity, metal affinity, ion exchange, hydrophobic interaction, and gel-filtration chromatography, have also been used to isolate membrane proteins. However, the proteins first must be released from the lipid environment with buffer solutions containing detergents. The choice of detergent plays an important role in extracting and purifying membrane proteins in an active form that is amenable to further downstream applications. A multitude of biochemical-grade detergents with various combinations of hydrophobic and hydrophilic groups are now commercially available for membrane protein isolation. They have been broadly classified on the basis of their polar "head" group as ionic, nonionic, or zwitterionic. Detergents have complex physicochemical properties which must be taken into consideration to be effectively employed when working with membrane proteins (Privé, 2007). These properties include the critical micelle concentration (CMC, the minimal concentration required for micelle formation), aggregation number, micelle molecular mass, hydrophilic–lipophilic balance (HLB), cloud point, and spectral properties. High-throughput multiwell-plate-based methods are available for rapid assessment of detergents for extraction and subsequent purification of membrane proteins; the identified detergents are then used for large-scale purification and crystallization (Gabrielsen et al., 2011; Gordon et al., 2008). When starting expression levels are very low, it is challenging to obtain membrane proteins of sufficient quantity for biochemical characterization and X-ray crystallographic structure elucidation. Tags and fusion partners such as His tag, FLAG tag, and green fluorescent protein (GFP) are useful not only for monitoring the expression and localization of membrane proteins but also for rapid purification of membrane proteins using affinity chromatographic methods (Fan et al., 2011; Rodríguez-Banqueri et al., 2012), as long as the tag does not affect the functional activity of the protein. Furthermore, it can be difficult to remove the fusion tag following purification of the protein. However, if the heterologously expressed protein is functionally/biochemically comparable to the plant-produced protein, irrespective of the presence of the tag, tagging may provide a method for the production/purification of the intractable protein.

One additional consideration for the isolation of heterologously produced membrane proteins concerns protein safety testing (e.g., acute oral toxicity testing in mice). It may be very difficult to identify detergents which can maintain the solubility of the isolated membrane protein, yet not confound the results of toxicity testing due to the potential inherent toxicity of the detergent. Furthermore, published data on the acute toxicity of detergents is lacking (Gad et al. 2006).

2.1.5.3. Activity assessment of membrane proteins. Most membrane proteins are not functional when removed from a lipid environment. Reconstitution of purified membrane proteins into liposomes is essential to reconstitute activity and enable biochemical studies, and a number of strategies have been used. Four different types of model membrane systems are reviewed by Shen et al. (2013). Reconstitution is particularly challenging when more than one protein is needed to obtain a functional complex, as it requires simultaneous expression of more than one membrane protein in addition to reconstituting the protein complex. Because of the challenges associated with liposome-based membrane protein reconstitution, nanodiscs are gaining more traction for the assembly of membrane proteins into phospholipid bilayers. The advantage of nanodiscs is that they create a native-like phospholipid bilayer that keeps membrane proteins in a soluble and functional form suitable for a number of downstream applications (reviewed by Bayburt and Sligar, 2010). Even then, co-factors may be needed, which requires the establishment of the entire system/pathway

in vivo. However, in many cases it is possible to assess protein activity by incubating the membrane fractions containing proteins of interest with labeled substrate and following the formation of enzymatic products. This method is especially suitable for membrane proteins involved in desaturation of fatty acids. A specific example was described by Madduri et al. (2012), who isolated microsome fractions from recombinant organisms expressing desaturases, and determined protein activity by incubating with non-natural substrates.

2.1.5.4. Detection and quantification of membrane proteins. Much of the protein, especially in the case of integral membrane proteins, can be buried within the membrane, limiting the exposed epitopes useful for immunological approaches. Even with alternative strategies (e.g., genetic immunization and antibody generation with synthetic peptides), it may be difficult or impossible with current methods to produce immunospecific antibodies because of the limited surface availability and poor antigenicity of the hydrophobic portions.

Selected- or multiple-reaction monitoring by tandem mass spectrometry, in a "bottom up" manner, was developed for absolute quantification (Houston et al., 2011; Stevenson et al., 2012) of proteins without using antibodies. Mass spectrometry–based alternative methods may offer an opportunity to detect or quantify some of the membrane proteins that have been intractable so far.

2.1.6. Example of an intractable membrane protein of agricultural importance

Diacylglycerol acyltransferase (DGAT) is a membrane protein associated with *in planta* high-oil traits. DGAT is the last enzyme in the triacylglycerol (TAG) synthesis pathway, and it functions to transfer an acyl group to the *sn*-3 position of 1,2-diacylglycerol to produce TAG (Ohlrogge and Browse, 1995). Over-expression of DGAT has been shown to increase seed oil content in maize and canola (reviewed by Shen et al., 2010; Oakes et al., 2011). Oil from DGAT-expressing high-oil crops plants has industrial (biodiesel) and nutritional applications (high oleic acid) (Zheng et al., 2008). DGATs are hydrophobic proteins found mainly in the endoplasmic reticulum (ER) and can contain up to ten putative transmembrane domains (Lung and Weselake, 2006; Shockey et al., 2006; Turchetto-Zolet et al., 2011; Yen et al., 2008). Zheng et al. (2008) predicted that the DGAT1-2 protein had eight transmembrane domains, while newer models predict nine transmembrane domains in the DGAT1-2 protein (Fig. 2 [Zhenglin Hou, DuPont Pioneer, personal communication]). DGATs are difficult to purify to homogeneity, and the use of detergents and chaotropic agents routinely used to maintain solubility often results in diminished or elimination of enzymatic activity (Lung and Weselake, 2006). Characterization studies with recombinant DGAT pose similar solubilization/purification challenges (Cao et al., 2011), thus creating additional challenges for hazard assessment tests.

The level of DGAT1-2 expression in Sf9 insect cells is 1–2 mg/L (DuPont Pioneer, unpublished results), which makes performing any toxicology testing of isolated protein very difficult or impossible using current methods.

There are no published examples of acute or repeated-dose toxicity studies for any integral membrane protein. Conventional heterologous over-expression systems (*E. coli*, insect cell, or yeast) have been used to produce recombinant DGAT (Cao et al., 2011; Lardizabal et al., 2001; Zheng et al., 2008); however, current capabilities do not allow for the isolation of active DGAT proteins from these systems in sufficient quantities for rodent studies (gram-level quantities).

Therefore, history of safe use concerns for DGAT can be addressed by analyzing and citing reports that identify DGAT isoforms in numerous field and vegetable crops, thus establishing prior dietary exposure in animal and humans (Constable et al., 2007; Section 3.2.1 of this manuscript).

A recent publication by Schroeder et al. (2013) describes the utility of various membrane transporters for improving sustainable food production. It provides examples for aluminum tolerance in acid soils, improved salt tolerance, pathogen resistance, solutions for iron and zinc deficiencies, and nitrogen uptake. These membrane transporters would most likely all be examples of intractable proteins.

2.2. Signaling proteins

2.2.1. Signaling and signal transduction proteins

These are a diverse group of proteins sharing the unique function of signaling in the presence of a metabolite, an environmental factor, or a cellular condition. This signaling in turn elicits a response, usually through a cascade of events, that results in multiple (pleiotropic) effects. This group includes protein/peptide hormones, environmental sensing/receptor proteins (modified in response to stress signals such as nitric oxide, carbon monoxide, salt, etc.), *R*-gene products (covered in more detail in Section 2.5), protein kinases/phosphatases, GTPases, defensins, and florigens (Bahyrycz and Konopińska, 2007; Batut et al., 2011; Bykova et al., 2011; Ghanem et al., 2011; Matsoukas et al., 2012; Mithoe and

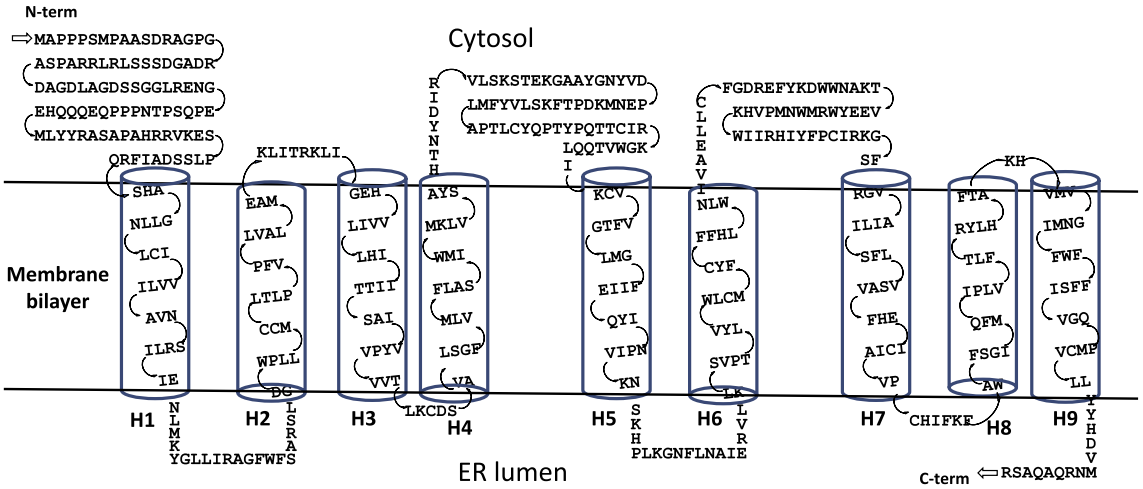

Fig. 2. Predicted structure of DGAT1-2.

Menke, 2011; Perilli et al., 2010; Rayapuram et al., 2012; Roberts, 2003; Spadaro et al., 2010).

2.2.2. Intractability issues associated with signaling proteins

Signaling pathways are often tightly controlled and regulated to avoid abnormal effects on plant growth and development. The expression levels of many protein kinases and protein phosphatases and other signaling proteins have been demonstrated to be very low in plant tissues. When transgenically expressing one or more such proteins in plants as a beneficial trait, the expression usually needs to be limited to a specific tissue (e.g., pollen, root, flower, endosperm, or embryo) and/or a certain developmental stage (Yang et al., 2012), and the expression level needs to be fine-tuned to minimize any potential adverse effects at the whole-plant level. In addition, there may be very few homologues in protein databases, and assessments made based on the safety of homologous proteins will be difficult to formulate.

An example illustrating the extremely low abundance of some proteins in this class is phosphoenolpyruvate carboxylase kinase (PEPCK). Tremendous efforts (e.g., Murmu and Plaxton, 2007; Saze et al., 2001; Wang and Chollet, 1993) have been undertaken to purify PEPCK from plants for biochemical characterization and molecular cloning. In one case (Saze et al., 2001), 0.1 μg of maize PEPCK was purified from 2.6 kg of leaves using a time-consuming and labor-intensive eight-step purification procedure. The purification factor was about 1.4 million fold.

2.3. Transcription factors

2.3.1. Definition of transcription factors

Transcription factors and their accessory proteins (TFAPs) regulate expression of a variety of genes in eukaryotic systems. Transcription factors bind to a specific DNA sequence and regulate transcription of DNA. A typical plant transcription factor consists of a DNA-binding region, an oligomerization site, a transcription regulation domain, and a nuclear localization signal (Liu et al., 1999). TFAPs generally do not bind DNA but interact with other proteins in the transcription complex to regulate gene transcription. Generally transcription factors and TFAPs regulate a family of genes and exert their influence by either up-regulating or down-regulating gene expression (Liu et al., 1999).

Transcription factors can be classified into different families based on their structural features. There are currently 84 gene families in the plant transcription factor database (Pérez-Rodríguez et al., 2010); examples include zinc finger transcription factors, bZIP transcription factors, Myb-related transcription factors, homeodomain transcription factors, Myc bHLH transcription factors, AP2 transcription factors, and E2F-DP transcription factors (De Veylder et al., 2007).

Transcription factors and TFAPs could play an important role in improving intrinsic yield potential and stability, abiotic stress tolerance, disease resistance, and nutrient use efficiency in plants (Century et al., 2008). For example, engineering with ZmNF-YB2 resulted in improved corn yields on water-limited soil (Nelson et al., 2007).

2.3.2. Intractability issues associated with transcription factors

2.3.2.1. In planta transcription factor expression levels. Transcription factors, such as the *Arabidopsis thaliana* B-box protein CONSTANS (CO), accumulate at levels too low to be detected by western blot analysis, even when functional activity is present (Suárez-López et al., 2001). Transcription of *CO* is known to be controlled by a number of factors, one of which is the circadian clock, which causes rhythmic oscillations in *CO* expression (Suárez-López et al., 2001). In addition, light regulates CO protein stability and activity (Imaizumi and Kay, 2006). Even for proteins that can be expressed in a heterologous system, enough functionally active protein must be isolated from the plant to test for equivalence, but it may be impossible with current methods to obtain sufficient quantities at the appropriate purity level to perform the typical characterization studies to establish equivalence of the *in planta* and heterologously expressed proteins.

2.3.2.2. Production of transcription factors in heterologous systems. Transcription factors produced in *E. coli* are often insoluble (Al-Samarrai et al., 2007; Yang et al., 2011; Zaret and Stevens, 1995). Obtaining even small quantities of the protein in a soluble, properly folded form requires refolding of the protein from inclusion bodies, which can be difficult or impossible with current methods (e.g., Al-Samarrai et al., 2007; Zaret and Stevens, 1995).

2.4. N-glycosylated proteins

2.4.1. Definition of N-glycosylated proteins

Intractable proteins include those proteins that are stably modified after translation, allowing for the "exponential diversification of the genome" (Webster and Thomas, 2012). Of all the possible post-translational modifications, N-glycosylation is the most common stable modification to impact the physicochemical properties of the protein. For some proteins, the type and extent of modification is specific to the plant species, the life stage, or the environment of the plant, or even to each organelle within the cell (see Rayon et al. (1998) for a review of N-glycosylation in plants).

N-glycosylation occurs in the secretory pathway and is well conserved in animals, plants, fungi, and social amoebae (Kukuruzinska and Lennon, 1998). Glycosylation in every species starts with a common pentasaccharide core sequence, but the similarity of the final structures varies. Plants have a different oligosaccharide makeup than do animals or fungi, and even within the plant kingdom, the sugars used and their positions in the substituted oligosaccharide vary among species. Heterogeneity of glycosylated proteins is observed at three different levels: the number of glycan side-chains, the extent of glycan modification of the different side-chains of the same glycoprotein, and the heterogeneity of oligosaccharide structures on the same N-glycosylation site (Lerouge et al., 1998).

While isolation and purification may be the overriding issue with some classes of intractable proteins, the more important issue for N-glycosylated proteins may be showing equivalence or comparability of the protein produced in the crop plant to that produced in a model system traditionally used to produce large amounts of protein. Even if significant amounts of protein can be produced in a heterologous eukaryotic system, the question arises whether the oligosaccharide substitution pattern will be equivalent, giving the protein the same folding properties, stability, and immunoreactivity as in the protein originating from the GM crop.

Glycosylation pattern can affect the folding and biological activity of a protein (Bosch and Schots, 2010). Burén et al. (2011) showed that CAH1 (chloroplast-localized carbonic anhydrase) is glycosylated at four sites (in some cases, five), and that glycosylation is necessary for correct folding, trafficking, and functionality of the protein. Conversely, the non-glycosylated protein formed aggregates and was retained in the ER and associated with ER chaperones, indicating that glycosylation of CAH1 is what facilitates folding and ER export. This result calls into question the validity of any safety assessment performed with naked CAH1 protein or any substituted glycoprotein. This also demonstrates a case where HOSU of a protein in one crop may not be applicable to a safety assessment in a different crop.

2.4.2. Intractability issues associated with post-translationally modified proteins

Given the effects that N-glycosylation can have on protein stability, immunogenicity, and/or activity of a protein, determining equivalence may be important, but the inability to determine the exact nature and composition of N-glycosylated proteins may be less important for any safety assessment as it should be noted that, to date, the only glycosylation known to result in adverse effects is the galactose-α-1,3-galactose that is present on some mammalian proteins and identical to tick protein glycans. It has been reported that some individuals bitten by ticks become sensitized to galactose-α-1,3-galactose and respond with allergic reactions following consumption of large amounts of meat or in other cases following exposure to the chemotherapeutic drug cetuximab (Berg et al., 2014; Commins and Platts-Mills, 2013). Though this does not indicate that every N-glycosylated protein should be tested for this particular moiety, it demonstrates that characterization of N-linked glycosylation is not likely to be helpful in the safety assessment process.

Glycosylation patterns vary from one organ to another in the plant and are also dependent on the developmental stage, leading to a population of proteins with different levels of glycan maturation. The importance of this maturation process to biological activity is an area of active research (Fanata et al., 2013). N-glycosylated proteins illustrate the difficult or even impossible situation that exists when trying to isolate and purify certain plant proteins produced in a heterologous system and then demonstrate their comparability/equivalence with the population of proteins produced by the GM crop. While not all proteins undergo irreversible post-translational modification and even fewer do so at the stoichiometric level that would be required for proper folding or function, such proteins present a unique challenge for safety assessment.

2.4.3. Challenges of using different heterologous expression systems for post-translationally modified proteins

Post-translational modifications in proteins limit the types of expression systems that can be used, particularly in the case of glycosylated proteins, since the glycosylation pattern can be unique to each organism. While *E. coli* lacks many of the post-translational modifications found in eukaryotes (Brondyk, 2009; Demain and Vaishnav, 2009), a recent publication (Valderrama-Rincon et al., 2012) shows that *E. coli* can be engineered to accomplish N-glycosylation and may offer opportunities in the future. Likewise, *P. pastoris* strains and baculovirus/insect cells have been engineered to introduce the human glycosylation machinery and to down-regulate their endogenous pathways (Hamilton et al., 2006; Harrison and Jarvis, 2006). However, their application for safety assessment of GM crops would require the introduction of the glycosylation pathway of each plant species to be considered, as the pattern or positioning of glycosylation in the expression system may not be the same as in the target crop. An example of this is alpha-amylase inhibitor-1 (αAI) presented later in this section. Even organ specificity may vary, ranging from the absence of organ specificity in the N-glycan processing of phytohemagglutinin (PHA) (Rayon et al., 1998) to considerable differences in relative abundance of N-glycans between different organs in *Brassica napus* (Gomord et al., 2010). Differences in the structure of the glycosyl part of a glycoprotein can also be seen at different developmental stages (Brooks, 2004). As a result, the determination of comparability between N-glycosylated proteins prepared in two different plants, let alone two different eukaryotes, may be difficult or impossible with current methods.

2.4.3.1. Example of an N-glycosylated protein. Alpha-amylase inhibitor-1 (αAI) from the common bean (*Phaseolus vulgaris*) demonstrates the issues of equivalence and multiple products seen with an N-glycosylated protein. Common bean αAI1 is synthesized as an inactive single-chain precursor that is proteolytically processed into α and β subunits, which then assemble into a mature αAI-1 molecule displaying inhibitory activity against insect α-amylases (Franco et al., 2002; Kluh et al., 2005). The two chains, α and β, form a heteromer and are glycosylated. The dominant form of the α-chain has both of its glycosylated sites occupied, while the β-chain's dominant form has only one of its two potential glycosylation sites occupied (Sawada et al., 2002; Yamaguchi et al., 1992; Young et al., 1999).

When the gene from common bean was expressed in pea (*Pisum sativum*) seed, the N-glycosylation pattern of the α-amylase inhibitor was different from that found in the bean seed (Marsh et al., 2011; Prescott et al., 2005). While Prescott suggested that the modified αAI protein in pea and not the native form was predisposed to cause antigen-specific inflammation, Lee et al. (2013) later showed that this allergenic response was not related to the αAI protein. Mass spectrometry identified a complex mixture of glycosylated αAI proteins. The α chains of αAI from common bean and GM pea showed the same glycosylation pattern, albeit in differing proportions, whereas the variation in the β chains was more complex, but still comparable. However, the difference in glycan pattern was smaller between the common bean and GM pea than between the common bean and two other bean species tested. Additionally, two studies (Campbell et al., 2011; Marsh et al., 2011) looked at the processing of αAI from the precursors to the mature α and β chains. Campbell et al. (2011) showed that the bean had more unprocessed or incompletely processed αAI precursors than the GM pea.

It is worth noting that based on the information described above, Marsh et al. (2011) concluded that "the transgenically expressed and native bean inhibitors can be considered to be 'substantially equivalent' within the wider context of variation within bean species that are consumed by humans".

2.5. Resistance (R)-proteins

2.5.1. Definition of R-proteins

Resistance (R) proteins in plants are involved in pathogen recognition and subsequent activation of innate immune responses (van Ooijen et al., 2008). There are two layers of defense in the plant immune system. The primary response is referred to as pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI). In PTI, the plant detects a common molecular pattern of a pathogen such as cell wall fragments, chitin, or peptide motifs in bacterial flagella as the pathogen tries to invade the plant. Recognition of the PAMP by the host plant receptor triggers the induction of the primary or basal defense responses (Tameling and Takken, 2008). To inhibit PTI, pathogens have evolved the ability to secrete effector proteins that suppress the basal plant defense response. In turn, plants have evolved means to recognize pathogen effectors and to mount a robust amplified defense response (effector-triggered immunity [ETI]). In the ETI response, the plant R-proteins detect the presence of pathogenic effector proteins (also referred to as avirulence factors) and trigger the defense response. The latter response is generally faster and stronger and is often accompanied by localized cell death at the infection site (hypersensitive response) (Jones and Dangl, 2006).

2.5.2. Subclasses of R-proteins

R-proteins display modular domain structures (Fig. 3). Based on these structural domains, the majority of R-proteins can be divided into four classes: two intracellular and two extracellular (van Ooijen et al., 2007). R-proteins in the two intracellular classes contain a central nucleotide-binding site (NBS) and a C-terminal leucine-rich repeat (LRR) domain (NBS-LRR proteins), but the two classes differ

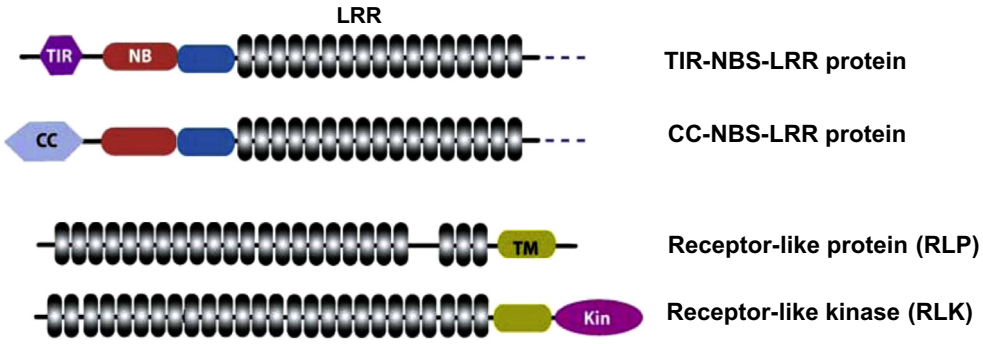

Fig. 3. Schematic representation of domains found in plant leucine-rich repeat (LRR) R-proteins. Domains are not drawn to scale. TIR, Toll/interleukin-1 receptor; CC, coiled-coil domain; NB or NBS, nucleotide-binding domain or site; LRR, leucine-rich repeat; TM, transmembrane; Kin, kinase. The linker region (dark blue) is not present in all proteins, e.g., Rpi-blb1. Figure adapted from Tameling and Takken (2008).

in their N-terminal domain. One class has a Toll and interleukin-1 receptor-like (TIR) domain at the N-terminus (TIR-NBS-LRR proteins) and the other carries a coiled-coil (CC) sequence (CC-NBS-LRR proteins). R-proteins in the two extracellular classes contain a predicted extracellular LRR (eLRR) domain at their N termini. This eLRR is connected via a transmembrane domain to a variable cytoplasmic C-terminal region. When the cytoplasmic domain contains a protein kinase domain, the R-protein is placed in the receptor-like kinase (RLK) class. If no such domain is present, it is placed in the receptor-like protein (RLP) class (van Ooijen et al., 2007).

2.5.3. Intractability issues associated with R-proteins

Most R-proteins are NBS-LRR proteins. The NBS proteins belong to the Signal Transduction ATPase with Numerous Domains (STAND) superfamily of ATPases. NBS-LRR genes have been identified in all plant species tested and thus seem to be a general mechanism of defense against pathogens (McHale et al., 2006).

Despite some key breakthroughs, progress toward understanding how R-proteins function has been slower than anticipated. Because of their low abundance in the plant cell (and in at least some cases rapid degradation upon effector recognition [Boyes et al., 1998]) and the lack of direct interaction between R-proteins and most of their corresponding effectors (multiprotein complex), many R-proteins have proven difficult to study biochemically (Luderer et al., 2001) and are not adaptable to functional assays.

2.5.4. Example of an R-protein

The *Rpi-blb1* gene from *Solanum bulbocastanum* encodes the Rpi-blb1 protein, which is a member of the CC-NBS-LRR class of R-proteins. Rpi-blb1 protein confers resistance to *Phytophthora infestans*, which causes potato late blight disease (van der Vossen et al., 2003). The *Rpi-blb1* gene has been introduced by genetic engineering into different potato cultivars in recent years (Vleeshouwers et al., 2008). In resistant plants, Rpi-blb1 protein detects the presence of effector protein Avrblb1 (IPI-O family class I) from *P. infestans* and triggers a host defense response.

Little information is available regarding Rpi-blb1 protein expression in GM crops. There appears to be low basal expression of most *R*-genes (Kramer et al., 2009). Despite major advances in understanding plant disease resistance mechanisms, it is still unclear exactly how Rpi-blb1 protein functions in the plant. No *in vitro* functional assay has been developed for Rpi-blb1 protein to date, so it will be difficult to confirm the functionality of the bacterially produced Rpi-blb1 protein.

Accordingly, it is not surprising that the *R*-gene product Rpi-blb1 was not detectable by standard ELISA such as typically used for measuring protein levels in transgenic plants (BASF Plant Science, unpublished results). This Rpi-blb1 ELISA has a lower limit of detection of approximately 100 ng/g dry weight of leaf material, corresponding to 100 ppb. Proteins encoded by *R*-genes are believed to be present at extremely low levels, and literature reports focus on transcript levels rather than protein levels (Kramer et al., 2009). If Rpi-blb1 is assumed to be present at only one copy per cell and we assume that there are approximately 100,000 cells/g leaf (calculated from 5000 cells/leaf [Autran et al., 2002] and 50 mg per Arabidopsis leaf), then there would be approximately 18 fg Rpi-blb1/g leaf (0.018 parts per trillion [ppt]; based on Rpi-blb1 MW of ~110,000 and 6.02×10^{23} molecules/mole). Even if this calculation is 1000-fold off, detecting 18 ppt is not within the current capabilities of today's technologies. Thus, Rpi-blb1 is clearly an example of an intractable protein due to its very low level of expression. The need to perform a comprehensive safety assessment is questionable because the exposure is virtually nonexistent. (If exposure is calculated based on ~200 kg of potato consumption/person/year [e.g., for Belarus in 2009; FAOSTAT, http://faostat.fao.org/site/609/DesktopDefault.aspx?PageID = 609#ancor] and 18 fg/g, the exposure is only 3.6 ng/year.) Furthermore, attempts to produce higher quantities of Rpi-blb1 in plant cells using constitutive high-level promoters were unsuccessful, perhaps due to the role the protein plays in protecting the plant from the pathogen.

3. Safety assessment of intractable proteins

The previous pages have described numerous classifications of intractable proteins and some of their possible applications in agricultural biotechnology. As a preface to any discussion about the safety of intractable proteins, it should be stressed that to date none of these classes of proteins has ever presented any evidence of hazard. Therefore, the mere use of a protein that would be considered intractable in a GM plant does not necessarily indicate that the expression of that protein presents any risk. Nevertheless, it is anticipated that regulatory authorities will request safety data for intractable proteins when used in GM plants. The following sections of this document will review how the components of protein safety assessment could be applied to intractable proteins.

If a protein is so unstable that it cannot be isolated, or it requires a membrane-type environment to be active, any degradation by digestion or processing would render the protein inactive; therefore, there may be little to gain, if anything, from any further safety testing. Hammond et al. (2013) recommended considering additional factors in protein safety assessment, including the impact of food processing on dietary exposure of introduced proteins, HOSU, and potential for harmful interactions between introduced proteins in crops containing multiple traits. For toxicology testing,

Hammond et al. (2013) recommended that a 28-day repeat-dose study should not be required if the introduced protein is structurally/functionally related to a family of proteins that has a HOSU based on bioinformatics and literature review, the protein is susceptible to inactivation during normal processing and cooking of foods, and the mode of action has been established (Hammond et al., 2013). These principles, built upon those recommended by Delaney et al. (2008), highlight that testing of proteins in animal toxicology studies may not be required for the safety assessment of GM proteins. These principles should be equally, if not more, applicable to intractable proteins.

However, if this rationale is not considered sufficient to predict the safety of an intractable protein, there are a number of tools that can be used. Many of these tools are adapted from well-established methods used today with "normal", tractable proteins, but some methods are new in that they offer a way to examine the properties of a protein in an indirect or novel way. These tools include HOSU (Section 3.2.1), bioinformatics (3.2.2), threshold of toxicological concern (3.2.3), use of substitutes and enriched fractions (3.2.7), mode of action and specificity (3.2.5), "targeted omics" (3.2.6), acute and repeated dosing (3.3.1), and *in vitro* assays (3.3.2). While no single property determined from the methods mentioned above may determine the safety of the protein, some combination of properties can provide a weight of evidence toward the safety of the protein being studied.

3.1. Safety evaluations for intractable proteins using the weight-of-evidence approach

The weight-of-evidence approach is designed to systematically evaluate and identify potential hazards of novel proteins by comparing their physical properties, sources, and sequences with those of known protein toxins or allergens (Table 3). Tier II testing may not be needed if the protein of interest has the following characteristics: (1) has a HOSU, (2) does not show structural similarity to known protein toxins or allergens, (3) has expression levels that can be measured to determine exposure, (4) is readily digestible, and (5) acts as intended (Delaney et al., 2008).

Since intractable proteins may not be able to be measured or subjected to the tests mentioned earlier, they present a unique challenge for safety assessment. Protein digestibility/stability, equivalency, expression level/exposure, and acute and repeated toxicity are difficult or impossible to measure for the reasons discussed in section 2. In the remaining part of this paper we discuss different strategies within the context of the weight-of-evidence testing approach as applied to intractable proteins.

3.2. Tier I testing

3.2.1. HOSU for intractable proteins

The weight-of-evidence approach currently used to assess the safety of a novel protein (Codex Alimentarius Commission, 2009) considers a variety of different factors. The safety assessment of any protein should begin with gathering historical information about prior human exposure or consumption, which is compiled into the HOSU. The degree to which HOSU can be established for a given protein can provide valuable information about whether further testing may be necessary to identify potential hazards associated with the protein (Delaney et al., 2008).

Important factors in establishing a HOSU include the following (Constable et al., 2007):

- The safety of the donor organism from which the protein was obtained
- Evidence of dietary or environmental exposure to the protein
- Consumption levels of the protein in different populations
- The results of any animal studies previously conducted with the protein and/or similar proteins
- Epidemiological studies

Establishing the safety of an introduced protein does not necessarily require that the exact protein has previously been contained in foods. Rather, safety can be established based on structural or functional similarity to proteins with a HOSU. Determining the degree of similarity required between an introduced protein and an existing protein with a HOSU is usually established through bioinformatics analyses on a case-by-case basis (Delaney et al., 2008).

Some information about the HOSU should be readily available regardless of whether a protein is simple to isolate and characterize or intractable. This information should include the safety of the donor organism and, if available, evidence related to dietary exposure to the protein from that source. What is particularly important in the case of intractable proteins is that this information can be obtained without having to isolate any purified protein.

3.2.2. Bioinformatics approaches for safety assessment of intractable proteins

At an early stage of the safety evaluation, bioinformatics analyses are conducted to compare the amino acid sequence of a candidate novel protein with known protein allergens and toxins (Delaney et al., 2008; Ladics et al., 2011). To evaluate the potential allergenicity of a candidate novel protein, the protein sequence is compared with sequences of known allergenic proteins that are maintained in a database at the Food Allergy Research and Resource Program (FARRP), University of Nebraska (http://www.allergenonline.org/). Potential homologies between the amino acid sequence of the candidate protein and those of proteins in the allergen database are evaluated using alignment algorithms (Pearson, 2000; Pearson and Lipman, 1988). Alignments can be reviewed for identities greater than or equal to the 35% threshold over 80 or more amino acids in conjunction with the E scores (Ladics et al., 2011). Based on Codex Alimentarius (2009), the amino acid sequence is also evaluated if there are eight contiguous identical amino acid matches to the sequences of allergenic

Table 3
Weight-of-evidence elements.

Element	Purpose
Tier I: Potential hazard identification	
History of safe use (HOSU)	Documents familiarity, already consumed in the existing food chain
Bioinformatics analysis/ homology searches	Flags structurally similar toxic, allergenic, pharmacological, or physiological active proteins
Protein levels in edible parts of the plant and dietary intakes	Used to establish exposure estimates
In vitro digestibility and heat stability	Determines likelihood that the GM protein will be denatured and/or degraded under physiological conditions or during cooking/processing
Mode of action and functional specificity	Provides evidence of biological function and specificity of the protein
Tier II: Hazard characterization	
Acute toxicity assessment	Assesses potential mammalian acute toxicity
Repeated-dose toxicity assessment	Examines broad range of possible adverse outcomes/effects over 28 days

Adapted from Delaney et al. (2008).

proteins in that database. However, more recent studies suggest that evaluation of eight contiguous identical amino acid matches to the sequences of allergenic proteins is not predictive and may lead to false positives (Herman et al., 2009; Ladics et al., 2011).

Bioinformatics tools are also used to compare the sequence of candidate novel proteins with those of other proteins to determine whether they are similar to known protein toxins. While this type of analysis can be useful in establishing similarity between candidate novel proteins and toxic proteins, there is currently no universally available database in which the sequences of toxic proteins are maintained and curated, nor is there a uniform standard against which similarity can be established. What would be very useful in practice is multiple sequence alignment coupled with secondary structure assignment and the predictions that could be made from it, particularly if they were combined with a threading program for 3-dimensional modeling. Nevertheless, bioinformatics is a powerful tool that can be used in identifying sequence similarity between novel proteins and proteins that have a documentable history of causing adverse effects. In a similar way, bioinformatics can be equally important in demonstrating similarity to proteins with a HOSU. As with HOSU, it is important to note that this information can be obtained without isolating the protein.

3.2.3. Expression levels in edible plant parts, dietary intake, and threshold of toxicological concern

Most intractable proteins are expressed at very low concentrations in plants and hence the dietary exposure of the protein to humans will be very low. Risk is a product of hazard and exposure, and exposure to these low-expressing intractable proteins will be minimal (see, e.g., section 2.5.4). Combining that with the likelihood that hazard (toxicity) data will not be available or will be limited, how can risk be assessed in this scenario?

A risk assessment strategy called the threshold of toxicological concern (TTC) is actively being discussed for chemicals. Application of the TTC concept is recommended for ranking and prioritizing risks from exposure to substances for which toxicity data are lacking or limited, but which may be present in food at low concentrations and for which exposure analysis can provide sound intake estimates. The TTC for chemicals is 0.15 μg/person/day or higher (up to 1800 μg/person/day), depending upon whether the chemical has any structural similarity with known carcinogens and on its toxic potency (Kroes et al., 2000, 2004). Chemicals are divided into three groups depending on the relatedness of their chemical structures to those of other chemicals that pose minimal safety concerns or, on the other hand, suggest the potential for toxicity (Kroes et al., 2000, 2004). Similarly, the TTC concept should be applicable in assessing the toxicological risk from dietary exposure of compounds produced in the edible parts of GM crops.

The main reason that the TTC concept has not been applied to allergenicity during safety assessment of proteins is that the thresholds of exposure for most food protein allergens have not yet been determined. The levels required for sensitization and induction are unknown for most allergens and a very small amount (at ppm level) can elicit an allergic reaction (Taylor et al., 2010). There is ongoing research in this area to determine the minimum allergen doses capable of inducing primary sensitization or elicitation, but more work has to be done to define the threshold values for most allergens (Kruizinga et al., 2008; Taylor et al., 2009).

In the case of intractable proteins, if the level of the protein is extremely low in the GM crop and the protein is not similar to known allergens, i.e., does not share significant amino acid sequence homology with a known allergen, the TTC concept could be applied. In addition to the low level of many intractable proteins, normal food processing can denature and inactivate the protein, which may further decrease the dietary exposure of active protein (Hammond and Jez, 2011).

Efforts have already been made to determine the TTC for proteins. Hammond and Cockburn (2008) estimated a TTC value for protein toxicity based on a no-observed-adverse-effect level (NOAEL) calculated from data available for 40 subchronic studies. The calculated NOAEL across these 40 subchronic studies was 249 mg/kg, which when divided by a 100-fold uncertainty factor provided a TTC of 2.49 mg/kg or 149 mg/adult person/day (assuming 60-kg body weight [BW]). Acute TTC based on 30 acute studies was calculated to be 17.9 mg/kg BW, or 1074 mg/adult person/day. Comparison of these threshold levels to actual dietary exposure levels assumes a "worst-case" scenario in which there is no loss of the introduced protein when the GM crop is processed into food. It might be possible to further develop the TTC concept based on additional protein toxicity data available to determine TTC values for proteins or structural classes of proteins.

Hammond and Cockburn (2008) illustrated how the TTC concept can be applied to proteins introduced into GM crops, using CP4 EPSPS (5-enol-pyruvylshikimate-3-phosphate synthase) protein as an example. CP4 EPSPS is present in grain from glyphosate-tolerant maize at a level of approximately 14 ppm (Hammond and Cockburn, 2008). EPSPS has a HOSU and CP4 EPSPS has no homology to known toxins or allergens (Harrison et al., 1996; Padgette et al., 1996). Making the most conservative assumption, i.e., that 100% of the corn consumed is derived from glyphosate-tolerant maize, the estimated intake of introduced protein is 0.004 mg/kg BW/day or 0.27 mg/person/day for CP4 EPSPS (Hammond and Cockburn, 2008). This exposure level is 600-fold below the calculated TTC for proteins of 2.49 mg/kg described above. If the effects of normal food processing are taken into account, the actual dietary exposure of CP4 EPSPS protein decreases by another 100-fold, which results in ~60,000-fold lower exposure than the chronic exposure TTC level (Hammond et al., 2013).

Hence, TTC could be applied to some of the proteins introduced into GM crops, especially intractable proteins, many of which are expressed at very low levels that might be reduced even further during food processing. If the dietary exposure level of a particular introduced protein is above the TTC, it does not mean that the protein is not safe for consumption; rather, it indicates that appropriate toxicity studies may be warranted to further assess the safety of the novel protein.

If the dietary exposure of the protein is below the TTC, then there is minimal toxicological concern associated with the protein. Adoption of the TTC concept in safety assessment of introduced proteins could help to avoid the use of large numbers of animals in toxicity studies. If protein dietary exposure of the introduced intractable protein is less than the TTC, there is a HOSU, and the bioinformatics search raises no concerns, then there is no scientific justification for further toxicological testing.

3.2.4. In vitro digestibility and heat stability using enriched protein preparations

The stability of the protein to gastric digestion is a factor considered when assessing its allergenicity, activity, or toxicity potential. A protein that is rapidly digestible by a protease (e.g., pepsin) is theorized to have less potential to exert adverse health effects when consumed than one that is less readily digested. An allergenic protein is assumed to be stable long enough to cross the mucosal membrane of the intestinal tract, where absorption occurs. The existence of significantly sized protease-resistant fragments, generally accepted as larger than approximately 3000–3500 MW (Bannon et al., 2002; Lack et al., 2002), after simulated gastric digestion suggests that further studies might be appropriate to assess whether the protein has the potential to be allergenic, e.g., sequential digestion in gastric fluid followed by intestinal fluid, comparison with known allergenic epitopes, heat stability,

etc. Heat stability studies address both potential allergenicity and toxicity potential. Both allergenicity and toxicity potential are evaluated (separately) using a weight-of-evidence approach of which digestibility and heat stability are just two parameters.

For intractable proteins that cannot be heterologously expressed and purified, digestibility could be monitored using an enriched preparation of protein and quantifying the absolute abundance of that protein using modern mass spectrometry–based approaches. For example, for a fatty acid–modifying enzyme that is localized to ER membrane, purified microsomes from the host plant that contain the recombinantly expressed protein could be tested for enrichment by using antibodies or mass spectrometry. This microsome preparation could then be subjected to an *in vitro* digestion assay simulating the human gut, and the abundance of this protein monitored post-digestion.

An excellent example of this approach is the preliminary safety assessment of the membrane-bound protein delta 9 desaturase (D9DS), recently reported by Madduri et al. (2012). The weight of evidence for this intractable protein included both heat and digestibility studies on enriched or microsomal preparations. These data along with HOSU, bioinformatics, and mode of action were used to support the human health safety of D9DS as expressed in oilseed crops.

Furthermore, when the level of an intractable protein is too low to detect within the transgenic plant, it is unlikely to be allergenic because allergenic proteins are generally expressed at a high percentage of the protein in the food (Lehrer et al. 2002).

3.2.5. Mode of action and specificity

The mode of action of a transgenic protein can be helpful in establishing the range of biological activity that is being introduced into a plant. In the case of intractable proteins, it is likely that traditional biochemical studies will not be available, but indirect methods can provide useful tools to establish the mode of action and functional specificity of intractable proteins. For example, bioinformatics can be used to make hypotheses for mode of action on the basis of primary sequence and structural homology to specific functional class of proteins.

Carbonell and Faulon (2010) developed a web-based method to predict promiscuity of enzymes using vector machine-trained molecular signature sequences. Su and Lee (2013) used a slightly different approach and developed a method to predict kinase-specific phosphorylation sites on protein three-dimensional structures using a support vector machine approach in conjunction with information on linear motifs and spatial amino acid composition.

The approaches outlined above could be used to demonstrate that the protein has a very narrow spectrum of activity and therefore its expression in a plant would be expected to have no activity other than the intended activity. Uncertainty about the safe use of any particular protein is reduced if the mode of action of the protein is demonstrated to be not deleterious to the human or animal consuming the protein.

3.2.6. Additional approaches to support the safety assessment of intractable protein traits

Directed or targeted "omics" (e.g., proteomics, metabolomics, transcriptomics) represent fast-evolving and powerful technologies which might also provide data useful for establishing the mode of action of newly introduced traits. These tools have their most value when employed during the discovery or trait identification phase of product development rather than during the safety assessment phase. The application of these methods will require *in planta* expression of the intractable protein, but they benefit from not requiring purification of the actual protein. Among the currently available "omics" approaches, transcriptomics offers the most complete view of the cognate biomolecule and could therefore be a good choice when little information is available about the mode of action of a given protein. With the advent of parallel short-sequencing approaches, an entire transcriptome can be cataloged quickly and inexpensively. While collecting these data is becoming cheaper and easier, organizing the morass of information is still arduous and it is in the interpretation of these data where significant costs begin to accrue. Furthermore, since transcript and protein expression levels are not always concordant (Hajduch et al., 2010), comparative proteomics may also be advisable to interrogate the plant system for safety (e.g., allergen equivalency) or mode of action. Thus, while transcriptomics technology can provide a measure of equivalency in perhaps the most global manner of any "omics" approach, it is merely one piece of the puzzle in a systems approach to approximate a plant.

The utility of any "omics" approaches in the context of safety assessment of GM crops is still emerging as baseline omics databases continue to be developed. As with any aspect of safety assessment, experiments need to be properly controlled and well-defined, as these technologies can lead to unsupported or worse, erroneous conclusions if improperly used or interpreted. For example, two plants grown in the same field under the same growing conditions will always show variation in analyte composition, which in the absence of biological context (i.e., genetic and environmental variation), makes equivalence assessments difficult and/or meaningless. Therefore, a targeted approach with a specific trait-connected application, such as the impact of an introduced enzyme on a specific metabolic pathway, may be more appropriate and useful.

3.2.7. Protein samples in support of safety assessment

3.2.7.1. Equivalence of plant-produced and substitute protein samples. Due to low expression of most transgenic proteins within the GM crop, safety studies are routinely conducted by using highly purified protein substitutes derived from heterologous expression systems. In equivalence studies, key parameters of the protein substitute are established to confirm biochemical and functional equivalence to the plant protein, and thus to confirm its suitability for such studies as digestibility and toxicological studies (Raybould et al., 2013). While this approach may not be applicable for many intractable proteins due to extremely low expression levels or the inability to assess functional activity (e.g., transcription factors [section 2.3.2] and most signaling proteins [section 2.2.2]), it should be recognized that bioinformatics and molecular analytical tools can be used to demonstrate that the protein expressed in the plant at a low level is in fact equivalent to the substitute prepared in another system. The combination of experimental and bioinformatics data can provide meaningful evidence to support the equivalence of the two proteins.

When the equivalence of the heterologously expressed protein cannot be experimentally determined for reasons described above, the use of heterologously expressed protein substitutes can provide valuable information in support of the safety assessment of the intractable protein. This is a case where bioinformatics can provide substantial evidence supporting the equivalence of the protein substitute. *In silico* molecular modeling that predicts similar folding or surface distribution of residues may also be a powerful tool to demonstrate the equivalence of the substitute protein. Many molecular modeling analyses have been shown to give reliable predictions, particularly when the structure of a closely related homologue is used as a starting point (Jaroszewski, 2009; Schlick et al., 2011).

3.2.7.2. Enriched protein samples derived from plant material. In the case where the protein substitute cannot be expressed in a heterologous system, enriched protein samples (i.e., concentrated and

desalted plant crude extracts of the GM crop) could provide valuable alternatives when directly applied in safety studies such as acute toxicity studies. Enriched protein samples containing a 40-fold increase of the transgenic protein have been reported (ANZFA, 2000), and although concentrations may be much lower than suggested by the Organisation for Economic Co-operation and Development (OECD) guidelines, these studies will provide a certain level of confidence in safety assessments because they employ the same plant protein to which humans, animals, and the environment will be exposed.

3.3. Tier II testing

In the event that data gaps are identified or that information is not conclusive during Tier I testing, individual studies contained within Tier II may be considered. Because of the nature of intractable proteins, it is almost certain that the study design will differ from those traditionally conducted with novel proteins. This may include modified *in vivo* tests or *in vitro* tests that may substitute for *in vivo* tests. However, this does not necessarily indicate that the results are any less applicable to the safety assessment of these proteins.

3.3.1. In vivo animal testing with intractable proteins

3.3.1.1. Acute study. Acute toxicity studies are often conducted in mice by a single high dose of 2000 mg/kg via oral gavage (OECD, 2001, designed for small molecules). An alternative to this maximal dose could be a dose based on at least a 100-fold margin of exposure (Hammond et al., 2013). Routinely, these studies are conducted with purified proteins. In the case of intractable proteins, a characterized enriched fraction or substitute protein may be used with a safety margin as high as possible (if 100-fold cannot be attained). After dosing, animals are observed daily for 14 days and changes in body weight and any clinical signs of adverse effects are recorded. At the end of the in-life phase, animals are euthanized and a gross necropsy is performed. This kind of study requires gram-level quantities of protein to dose multiple mice. Although seldom used, intravenous (IV) administration of the test substance has been reported as an indication of a lack of toxicity of the phosphinothricin acetyltransferase (PAT) protein (Hérouet et al., 2005) and may be an alternative when only a small amount of protein is available for the acute study. The IV route of exposure provides high systemic exposure to the active test protein by avoiding denaturation in low gastric pH, the proteolytic degradation in the gut, and absorption kinetics, and in one way mimics a worst-case scenario; however, the protein bypasses the gastrointestinal (GI) system and therefore avoids a potential target of the protein. Also, when protein is produced in a heterologous bacterial system, a high level of bacterial endotoxin in the protein extract can lead to false positive results. Therefore, IV testing should be used with some caution.

3.3.1.2. Repeated-dose study. The purpose of the repeated-dose dietary study is to evaluate the test substance's propensity to elicit adverse effects in rodents up to 28 days after administration. Commonly done for small molecules, laboratory rodents (mice or rats) are repeatedly exposed to a limit dose of 1000 mg/kg of body weight of test substance (in accordance with OECD 407 [OECD, 1995]; see Delaney, 2007; Mathesius et al., 2009). Doses have also been administered at levels of up to 1000 times the estimated human exposure (Juberg et al., 2009), but such studies require knowledge of the concentration of the GM protein that is expressed in the edible part(s) of the GM crop. The endpoints that are typically measured in repeated-dose studies are more comprehensive than those in acute toxicology studies and include clinical and ophthalmological observations, assessment of body weights, feed consumption, functional observational batteries, motor activity, clinical chemistry, organ weights, and histopathology (Delaney, 2007).

Repeated-dose toxicology studies with GM proteins at the limit dose levels require > 20 grams of protein test substance (Delaney et al., 2008; Juberg et al., 2009; Mathesius et al., 2009), an amount unlikely to be obtained with intractable proteins. If it is determined, following Tier I testing, that *in vivo* toxicology studies are appropriate, it may be appropriate to test with enriched preparations containing the recombinant protein of interest. The caveat is that animal toxicology studies with enriched preparations would require that analytical methodologies be available to determine the concentration of the intractable protein in the preparation for purposes of dose extrapolation to estimated human exposure.

Microsome preparations, as referenced in section 3.2.4 for digestibility studies, could likewise be useful in *in vivo* toxicology studies. The use of microsomes for rodent toxicology studies would represent an excellent delivery system for integral membrane proteins, but it should be noted that microsomes represent a complex mixture of lipids and proteins (Han et al., 2001) and the amount of specific membrane protein could be extremely low. Nevertheless, if the concentration of the transmembrane protein in the enriched fraction is known and can be related to human exposure to determine margin of exposure, the results should be valid regardless of the presence of other ingredients in the test substance, particularly in the case that no adverse effects are observed.

The same considerations hold for using yeast as an expression and delivery system for *in vivo* toxicology studies. See Blanquet-Diot et al. (2007) for the use of yeast in biodrug development and Powilleit et al. (2007) for use of yeast in vaccine development.

Finally, it should be considered that feeding studies with whole foods or processed fractions from GM crops may have application in establishing a margin of exposure through evaluation of the crop containing intractable protein(s). While the limitations and caveats described in Bartholomaeus et al. (2013) and Kuiper et al. (2013) should be considered, this situation could provide information relating to exposure to the intractable protein(s) in a context that is more realistic under real-world conditions. In most cases, these types of studies would be expected to provide a good margin of exposure relative to humans as the concentrations at which the food and feed fractions are incorporated are well in excess of human exposure.

3.3.2. In vitro alternatives to in vivo testing

Most dietary proteins do not cause adverse effects to humans because they are metabolized into amino acids and small peptides in the gut that are readily absorbed for nutritive purposes (Delaney et al., 2008). Conversely, proteins that are known to be toxic to humans and other animals following oral exposure are often resistant to digestion by the target organism such that they have either toxicity primarily directed toward intestinal cells (e.g., kidney bean lectin [Lafont et al., 1988; Rossi et al., 1984; Weinman et al., 1989]) or can also be accompanied by systemic toxicity following absorption from the GI system (e.g., ricin [Cook et al., 2006; Ishiguro et al., 1992a, 1992b]). Furthermore, the potential toxicity of proteins introduced into GM crops has been evaluated in mice or other laboratory animals by exposure to purified proteins, though to date, none have demonstrated any evidence of adverse effects (for example, see Hérouet et al., 2005; Juberg et al., 2009; Mathesius et al., 2009; Rice et al., 2008).

Using an *in vitro* model of human intestinal epithelial cells to investigate the toxicity of proteins as an alternative to *in vivo* studies with animals may have merit particularly in the case of intractable proteins when isolation of gram quantities will not be possible. This model includes many of the structural features found in the human gastrointestinal system including development of

Table 4
Potential hazard identification tasks for intractable proteins.

Task	Is purified protein required?	Is conventional approach possible with intractable protein?	Solutions for testing intractable protein	Classes of intractable proteins for which solution is applicable	Limitations of test/approach
History of safe use (HOSU)	No	Yes	Systematic review of literature and a summary paper	All	Data gaps Conflicting results to be interpreted (subjective)
Bioinformatics analysis	No	Yes	*In silico* search with updated databases	All	False positives may lead to erroneous conclusions False negatives cannot be detected
Mode of action	Yes/No	No	Use structurally similar proteins with well-established mode of action to build hypothesis Directed "omics" approaches	Most useful for signaling proteins, transcription factors, R-proteins; could be used to supplement data in any category	Lacks direct experimental data Need for validation of "omics" data
In vitro digestibility and heat stability	Yes	No	Test enriched fractions / grain or crop protein extracts containing protein of interest	All	Difficulty in detection and limitations of identifying/characterizing digest fragments in complex mixture
Expression level and dietary intake	Protein in grain/crop	Yes/No	Proteomics approach and alternative techniques to measure protein levels	All	Need to define a robust technique to measure the level of protein present in plant tissue
Protein equivalence	Yes	No	Isolate protein from a transient or model plant expression system with alternative promoters/vectors	All, particularly signaling protein and transcription factors	Plant protein not isolated from the product event

Table 5
Potential hazard characterization tasks for intractable proteins.

Task	Is purified protein required?	Is conventional approach possible with intractable protein?	Solution for testing intractable protein	Classes of intractable proteins for which solution is applicable	Limitations of test/approach
Acute toxicity study (14-day)	Yes	No Limit dose = 2000 mg/kg bw/day[a] oral toxicity study	Dose with enriched prep /plant or recombinant protein extracts Proper control selection MOE[b] approach	All for which accurate quantitation can be made	Limit dose will not be achieved Testing complex mixture
Repeated-dose toxicity study (28-day)	Yes	No Limit dose = 1000 mg/kg bw/day	Dose with enriched prep/grain or crop protein extracts Proper control selection MOE approach	All for which accurate quantitation can be made	Limit dose will not be achieved Testing complex mixture
In vitro testing	Yes	Yes	*In vitro* cell-based cytotoxicity Proven robustness and predictability of the test needed	Most useful for signaling proteins, transcription factors, and R-proteins where quantities are limited; could be used for non-hydrophobic proteins in any category	Translational argument needs to be made (cell to whole organism) Not amenable to hydrophobic proteins

[a] bw, body weight.
[b] MOE, margin of exposure.

microvilli and formation of tight junctions. In this type of model, disruption in the integrity of intestinal cell monolayers following exposure to a protein could be considered as a sign of toxicity associated with that protein in the presence or absence of direct cytotoxicity.

The use of cultured intestinal epithelial cells in the safety assessment process is not entirely new, as they have been used to predict the bioavailability of drugs with a strong correlation to *in vivo* bioavailability studies (Artursson et al., 2012). In addition, *in vitro* toxicology using human-derived cells and tissues is emerging as an alternative to *in vivo* animal models and has gained acceptance in some areas of consumer product testing (e.g., cosmetics). The benefits of using *in vitro* models for conducting toxicology testing of GM proteins would include a reduction in the number of animals used in research, substantially less isolated protein required, high reproducibility, and lower cost.

While there are advantages to *in vitro* studies, there are limitations as well, including the fact that they cannot model the complexity of the gastrointestinal tract. Though *in vitro* studies have been accepted as a model for predicting drug bioavailability, they have not been validated with universally accepted experimental endpoints to define an adverse effect. Additionally, it is not currently understood how the concentrations of proteins used *in vitro* compare with doses of proteins that would be administered *in vivo*. In the absence of a validated *in vitro* model, there can be many confounding factors that can alter the outcome of the study. These factors can either affect the cells in culture (i.e., pH, osmolarity, nature and content of media, etc.) or be associated with the test protein (hydrophilic vs. hydrophobic, purity, etc.), which may confound the interpretation of the intrinsic toxicological effects of the test protein. Finally, this type of modeling may also require special attention to the positive and negative control substances to be used experimentally to demonstrate sensitivity and selectivity.

For these reasons, it is not yet known whether an *in vitro* toxicity test with cultured intestinal epithelial cells can be used as a replacement for mammalian toxicity studies, but evidence suggests that it should continue to be investigated and remains of considerable interest. Future investigations in this area will likely focus on determining whether the results obtained from *in vitro* studies are similar to those observed *in vivo*. Determining which variables to measure and which controls to use will be important considerations.

4. Conclusions

The potential benefits of intractable proteins include a broad range of valuable traits such as disease resistance, drought tolerance, nitrogen use efficiency, and enhanced nutrient value. Five categories of intractable proteins were identified and described here: membrane proteins; signaling proteins, transcription factors, N-glycosylated proteins, and R-proteins. Although each category has its specific challenges (Table 1), those common to several are (1) the very low concentrations of protein produced *in planta* and (2) the inability to obtain protein from a heterologous system that is (or can be adequately demonstrated to be) equivalent to the plant-produced protein. With the possible exception of N-glycosylated proteins, many of the same qualities (e.g., low stability and low quantity) that make these proteins intractable are the same qualities that indicate lack of hazards in conventional proteins. Thus the science indicates that thorough safety reviews are not necessary.

However, the classification of a protein as intractable does not preclude a comprehensive safety assessment of the protein should one become necessary. The tiered weight-of-evidence approach published by ILSI in 2008 (Delaney et al., 2008) and the modifications proposed by Hammond et al. (2013) propose that the type and extent of testing be tailored based on exposure levels of the identified hazards, rather than applied uniformly to every situation. This approach is particularly important in the case of intractable proteins, for which some of the conventional hazard identification and characterization tasks are very difficult or impossible with current technology.

The HOSU and bioinformatics analysis can be completed without isolation of any protein, and can thus be performed regardless of other challenges (Table 4). HOSU reflects knowledge both of the donor organism and of the protein itself, or of close relatives, which may have already been consumed by humans with no adverse effects. Bioinformatics analysis helps to identify which information on other proteins or species might be applicable to the intractable protein, as well as whether the protein itself has characteristics indicative of potential toxicity or allergenicity.

In addition, bioinformatics could provide helpful information when establishing the mode of action, especially when the nature of the protein has made it difficult to study directly. The remaining task in hazard identification, assessment of *in vitro* digestibility, has typically been performed with pure protein. An alternative to this approach is to use enriched proteins (e.g., microsome or similar preparations) from the GM crop or recombinant expression system. For integral membrane proteins, this approach has the advantage that the protein could be tested in its native form. Likewise, for those cases in which Tier II testing is judged to be necessary, the acute/repeated-dose tasks could be completed with semi-pure preparations (Table 5). In cases where enriched samples are used for feeding, it is necessary to be able to accurately quantitate the level of protein in the sample.

5. Conflict of interest statement

The employment affiliation of the authors is shown on the first page. The paper was prepared during the normal course of the authors' employment. Authors Bannon, Bushey, Delaney, Graser, Harper, Jiang, Lee, Madduri, Privalle, Ranjan, Schafer, and Zhang were employed at the time of authorship by companies that research and market genetically modified crops.

Preparation of this manuscript was financially supported by members of the ILSI International Food Biotechnology Committee: BASF Corporation; Bayer CropScience; Dow AgroSciences; Monsanto Company; Pioneer Hi-Bred, a DuPont business; and Syngenta Biotechnology, Inc.

Acknowledgments

This document has been reviewed by individuals internationally recognized for their diverse perspectives and technical expertise. However, it must be emphasized that the content of this document is the authors' responsibility and not the reviewers', and it does not represent an endorsement by the reviewers' institutions. While the authors carefully considered the views and comments of the various reviewers and the sponsors, the final analyses, interpretations and conclusions drawn are exclusively those of the authors.

The authors thank the following individuals for participating in the review process and for providing many constructive comments and suggestions: Reynaldo Ariel Alvares-Morales, Comisión Intersecretarial de Bioseguridad de los Organismos Genéticamente Modificados (CIBIOGEM), Mexico; Bronwyn Barkla, Universidad Nacional Autónoma de México; Rebecca Boston, North Carolina State University; Paul Brent, Food Standards Australia New Zealand (FSANZ); Steve Gendel, US Food and Drug Administration/Center for Food Safety and Applied Nutrition (FDA/CFSAN); Paul Keese,

Office of the Gene Technology Regulator (OGTR), Australia; Charlotte Madsen, Technical University of Denmark; Wayne Parrott, University of Georgia; Guri Rao, Iowa State University; and Rita Ulloa, Instituto de Investigaciones en Ingeniería Genética y Biología Molecular (INGEBI), Argentina.

The authors also wish to thank Kate Walker, former ILSI Scientific Program Manager, for her efforts in seeing this project to completion. The authors acknowledge the assistance of Virginia M. Peschke (Oakside Editorial Services) in the preparation of this paper.

References

Al-Samarrai, T.H., Kirk, C.A., Jones, W.T., Harvey, D., Sun, X., 2007. Expression in *Escherichia coli* and in vitro refolding of the plant transcription factor *Arabidopsis thaliana* RGL3. Protein Exp. Purif. 53, 289–292.

André, N., Cherouati, N., Prual, C., Steffan, T., Zeder-Lutz, G., Magnin, T., Pattus, F., Michel, H., Wagner, R., Reinhart, C., 2006. Enhancing functional production of G protein-coupled receptors in *Pichia pastoris* to levels required for structural studies via a single expression screen. Protein Sci. 15, 1115–1126.

Artursson, P., Palm, K., Luthman, K., 2012. Caco-2 monolayers in experimental and theoretical predictions of drug transport. Adv. Drug Deliv. Rev. 64 suppl., 280–289.

Asada, H., Uemura, T., Yurugi-Kobayashi, T., Shiroishi, M., Shimamura, T., Tsujimoto, H., Ito, K., Sugawara, T., Nakane, T., Nomura, N., Murata, T., Haga, T., Iwata, S., Kobayashi, T., 2011. Evaluation of the *Pichia pastoris* expression system for the production of GPCRs for structural analysis. Microb. Cell Fact. 10, 24.

Australia New Zealand Food Authority (ANZFA), 2000. Draft risk analysis report. Application A385: Food derived from insect-protected Bt-176 corn. 4 October 2000. Australia New Zealand Food Authority, Canberra, Australia. http://www.foodstandards.gov.au/code/applications/documents/A385%20FA.pdf. Retrieved 15 July 2013.

Autran, D., Jonak, C., Belcram, K., Beemster, G.T.S., Kronenberger, J., Grandjean, O., Inzé, D., Trass, J., 2002. Cell numbers and leaf development in *Arabidopsis*: a functional analysis of the *STRUWWELPETER* gene. EMBO J. 21, 6036–6049.

Bahyrycz, A., Konopińska, D., 2007. Plant signalling peptides: some recent developments. J. Peptide Sci. 13, 787–797.

Bannon, G.A., Goodman, R.E., Leach, J.N., Rice, E., Fuchs, R.L., Astwood, J.D., 2002. Digestive stability in the context of assessing the potential allergenicity of food proteins. Comments Toxicol. 8, 271–285.

Bartholomaeus, A., Parrott, W., Bondy, G., Walker, K., on behalf of the ILSI International Food Biotechnology Committee Task Force on the Use of Mammalian Toxicology Studies in the Safety Assessment of GM Foods. 2013. The use of whole food animal studies in the safety assessment of genetically modified crops: Limitations and recommendations. Crit. Rev. Toxicol. 43, 1–24.

Batut, J., Mergaert, P., Masson-Boivin, C., 2011. Peptide signalling in the rhizobium–legume symbiosis. Curr. Opin. Microbiol. 14, 181–187.

Bayburt, T.H., Sligar, S.G., 2010. Membrane protein assembly into Nanodiscs. FEBS Lett. 584, 1721–1727.

Berg, E.A., Platts-Mills, T.A.E., Commins, S.P., 2014. Drug allergens and food—the cetuximab and galactose-α-1,3-galactose story. Ann. Allergy Asthma Immunol. 112, 97–101.

Bernaudat, F., Frelet-Barrand, A., Pochon, N., Dementin, S., Hivin, P., Boutigny, S., Rioux, J.B., Salvi, D., Seigneurin-Berny, D., Richaud, P., Joyard, J., Pignol, D., Sabaty, M., Desnos, T., Pebay-Peyroula, E., Darrouzet, E., Vernet, T., Rolland, N., 2011. Heterologous expression of membrane proteins: choosing the appropriate host. PLoS ONE 6, e29191.

Blanquet-Diot, S., Garrait, G., Jarrige, J.F., Alric, M., 2007. Genetically modified yeasts as a new oral drug delivery system: from in vitro to in vivo validation of the concept. In: Méndez-Vilas, A. (Ed.), Communicating Current Research and Educational Topics and Trends in Applied Microbiology. Formatex, Badajoz, Spain, pp. 378–390.

Bosch, D., Schots, A., 2010. Plant glycans: friend or foe in vaccine development? Expert Rev. Vaccines 9, 835–842.

Bowie, J.U., 2005. Solving the membrane protein folding problem. Nature 438, 581–589.

Boyes, D.C., Nam, J., Dangl, J.L., 1998. The *Arabidopsis thaliana RPM1* disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. Proc. Natl. Acad. Sci. USA 95, 15849–15854.

Brondyk, W.H., 2009. Selecting an appropriate method for expressing a recombinant protein. Methods Enzymol. 463, 131–147.

Brooks, S.A., 2004. Appropriate glycosylation of recombinant proteins for human use: Implications of choice of expression system. Mol. Biotechnol. 28, 241–255.

Burén, S., Ortega-Villasante, C., Blanco-Rivero, A., Martínez-Bernardini, A., Shutova, T., Shevela, D., Messinger, J., Bako, L., Villarejo, A., Samuelsson, G., 2011. Importance of post-translational modifications for functionality of a chloroplast-localized carbonic anhydrase (CAH1) in *Arabidopsis thaliana*. PLoS ONE 6, e21021.

Bykova, N.V., Hoehn, B., Rampitsch, C., Banks, T., Stebbing, J.A., Fan, T., Knox, R., 2011. Redox-sensitive proteome and antioxidant strategies in wheat seed dormancy control. Proteomics 11, 865–882.

Campbell, P.M., Reiner, D., Moore, A.E., Lee, R.Y., Epstein, M.M., Higgins, T.J.V., 2011. Comparison of the α-amylase inhibitor-1 from common bean (*Phaseolus vulgaris*) varieties and transgenic expression in other legumes—post-translational modifications and immunogenicity. J. Agric. Food Chem. 59, 6047–6054.

Cao, H., Chapital, D.C., Shockey, J.M., Klasson, K.T., 2011. Expression of tung tree diacylglycerol acyltransferase 1 in *E. coli*. BMC Biotechnol. 11, 73.

Carbonell, P., Faulon, J.L., 2010. Molecular signatures-based prediction of enzyme promiscuity. Bioinformatics 26, 2012–2019.

Century, K., Reuber, T.L., Ratcliffe, O.J., 2008. Regulating the regulators: the future prospects for transcription-factor-based agricultural biotechnology products. Plant Physiol. 147, 20–29.

Charollais, J., Van Der Goot, F.G., 2009. Palmitoylation of membrane proteins. Mol. Membr. Biol. 26, 55–66.

Codex Alimentarius Commission, 2009. Foods Derived From Modern Biotechnology, second ed. FAO/WHO, Rome.

Commins, S.P., Platts-Mills, T.A.E., 2013. Tick bites and red meat allergy. Curr. Opin. Allergy Clin. Immunol. 13, 354–359.

Constable, A., Jonas, D., Cockburn, A., Davi, A., Edwards, G., Hepburn, P., Herouet-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 Chem. Toxicol. 45, 2513–2525.

Cook, D.L., David, J., Griffiths, G.D., 2006. Retrospective identification of ricin in animal tissues following administration by pulmonary and oral routes. Toxicology 223, 61–70.

Delaney, B., 2007. Strategies to evaluate the safety of bioengineered foods. Int. J. Toxicol. 26, 389–399.

Delaney, B., Astwood, J.D., Cunny, H., Eichen Conn, R., Herouet-Guicheney, C., MacIntosh, S., Meyer, L.S., Privalle, L., Gao, Y., Mattsson, J., Levine, M., and the ILSI International Food Biotechnology Committee Task Force on Protein Safety, 2008. Evaluation of protein safety in the context of agricultural biotechnology. Food Chem. Toxicol. 46 Suppl 2, S71–S97.

Demain, A.L., Vaishnav, P., 2009. Production of recombinant proteins by microbes and higher organisms. Biotechnol. Adv. 27, 297–306.

De Veylder, L., Beeckman, T., Inzé, D., 2007. The ins and outs of the plant cell cycle. Nat. Rev. Mol. Cell Biol. 8, 655–665.

Fan, J., Heng, J., Dai, S., Shaw, N., Zhou, B., Huang, B., He, Z., Wang, Y., Jiang, T., Li, X., Liu, Z., Wang, X., Zhang, X.C., 2011. An efficient strategy for high throughput screening of recombinant integral membrane protein expression and stability. Protein Expr. Purif. 78, 6–13.

Fanata, W.I.D., Lee, K.H., Son, B.H., Yoo, J.Y., Harmoko, R., Ko, K.S., Ramasamy, N.K., Kim, K.H., Oh, D.B., Jung, H.S., Kim, J.Y., Lee, S.Y., Lee, K.O., 2013. N-glycan maturation is crucial for cytokinin-mediated development and cellulose synthesis in *Oryza sativa*. Plant J. 73, 966–979.

Franco, O.L., Rigden, D.J., Melo, F.R., Grossi-de-Sá, M.F., 2002. Plant α-amylase inhibitors and their interaction with insect α-amylases: Structure, function and potential for crop protection. Eur. J. Biochem. 269, 397–412.

Gabrielsen, M., Kroner, F., Black, I., Isaacs, N.W., Roe, A.J., McLuskey, K., 2011. High-throughput identification of purification conditions leads to preliminary crystallization conditions for three inner membrane proteins. Mol. Membr. Biol. 28, 445–453.

Gad, S.C., Cassidy, C.D., Aubert, N., Spainhour, B., Robbe, H., 2006. Nonclinical vehicle use in studies by multiple routes in multiple species. Int. J. Toxicol. 25, 499–521.

Ghanem, M.E., Hichri, I., Smigocki, A.C., Albacete, A., Fauconnier, M.L., Diatloff, E., Martinez-Andujar, C., Lutts, S., Dodd, I.C., Pérez-Alfocea, F., 2011. Root-targeted biotechnology to mediate hormonal signalling and improve crop stress tolerance. Plant Cell Rep. 30, 807–823.

Gomord, V., Fitchette, A.C., Menu-Bouaouiche, L., Saint-Jore-Dupas, C., Plasson, C., Michaud, D., Faye, L., 2010. Plant-specific glycosylation patterns in the context of therapeutic protein production. Plant Biotechnol. J. 8, 564–587.

Gordon, E., Horsefield, R., Swarts, H.G.P., de Pont, J.J.H.H.M., Neutze, R., Snijder, A., 2008. Effective high-throughput overproduction of membrane proteins in *Escherichia coli*. Protein Expr. Purif. 62, 1–8.

Hajduch, M., Hearne, L.B., Miernyk, J.A., Casteel, J.E., Joshi, T., Agrawal, G.K., Song, Z., Zhou, M., Xu, D., Thelen, J.J., 2010. Systems analysis of seed filling in Arabidopsis: using general linear modeling to assess concordance of transcript and protein expression. Plant Physiol. 152, 2078–2087.

Hamilton, S.R., Davidson, R.C., Sethuraman, N., Nett, J.H., Jiang, Y., Rios, S., Bobrowicz, P., Stadheim, T.A., Li, H., Choi, B.K., Hopkins, D., Wischnewski, H., Roser, J., Mitchell, T., Strawbridge, R.R., Hoopes, J., Wildt, S., Gerngross, T.U., 2006. Humanization of yeast to produce complex terminally sialylated glycoproteins. Science 313, 1441–1443.

Hammond, B.G., 2008. Food Safety of Proteins in Agricultural Biotechnology. CRC Press, Boca Raton, FL.

Hammond, B., Cockburn, A., 2008. The safety assessment of proteins introduced into crops developed through agricultural biotechnology: a consolidated approach to meet current and future needs. In: Hammond, B.G. (Ed.), Food Safety of Proteins in Agricultural Biotechnology. CRC Press, Boca Raton, FL, pp. 259–288.

Hammond, B.G., Jez, J.M., 2011. Impact of food processing on the safety assessment for proteins introduced into biotechnology-derived soybean and corn crops. Food Chem. Toxicol. 49, 711–721.

Hammond, B., Kough, J., Herouet-Guicheney, C., Jez, J.M., on behalf of the ILSI International Food Biotechnology Committee Task Force on the Use of Mammalian Toxicology Studies in the Safety Assessment of GM Foods, 2013.

Toxicological evaluation of proteins introduced into food crops. Crit. Rev. Toxicol. 43, 25–42.
Han, D.K., Eng, J., Zhou, H., Aebersold, R., 2001. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nat. Biotechnol. 19, 946–951.
Harrison, L.A., Bailey, M.R., Naylor, M.W., Ream, J.E., Hammond, B.G., Nida, D.L., Burnette, B.L., Nickson, T.E., Mitsky, T.A., Taylor, M.L., Fuchs, R.L., Padgette, S.R., 1996. The expressed protein in glyphosate-tolerant soybean, 5-enolpyruvylshikimate-3-phosphate synthase from *Agrobacterium* sp. strain CP4, is rapidly digested in vitro and is not toxic to acutely gavaged mice. J. Nutr. 126, 728–740.
Harrison, R.L., Jarvis, D.L., 2006. Protein *N*-glycosylation in the baculovirus–insect cell expression system and engineering of insect cells to produce "mammalianized" recombinant glycoproteins. Adv. Virus Res. 68, 159–191.
Henningsen, R., Gale, B.L., Straub, K.M., DeNagel, D.C., 2002. Application of zwitterionic detergents to the solubilization of integral membrane proteins for two-dimensional gel electrophoresis and mass spectrometry. Proteomics 2, 1479–1488.
Herman, R.A., Song, P., ThirumalaiswamySekhar, A., 2009. Value of eight-amino-acid matches in predicting the allergenicity status of proteins: an empirical bioinformatic investigation. Clin. Mol. Allergy 7, 9.
Hérouet, C., Esdaile, D.J., Mallyon, B.A., Debruyne, E., Schulz, A., Currier, T., Hendrickx, K., van der Klis, R.J., Rouan, D., 2005. Safety evaluation of the phosphinothricin acetyltransferase proteins encoded by the *pat* and *bar* sequences that confer tolerance to glufosinate-ammonium herbicide in transgenic plants. Regul. Toxicol. Pharmacol. 41, 134–149.
Houston, N.L., Lee, D.G., Stevenson, S.E., Ladics, G.S., Bannon, G.A., McClain, S., Privalle, L., Stagg, N., Herouet-Guicheney, C., MacIntosh, S.C., Thelen, J.J., 2011. Quantitation of soybean allergens using tandem mass spectrometry. J. Proteome Res. 10, 763–773.
Imaizumi, T., Kay, S.A., 2006. Photoperiodic control of flowering: not only by coincidence. Trends Plant Sci. 11, 550–558; erratum p. 567.
Ishiguro, M., Matori, Y., Tanabe, S., Kawase, Y., Sekine, I., Sakakibara, R., 1992a. Biochemical studies on oral toxicity of ricin. V. The role of lectin activity in the intestinal absorption of ricin. Chem. Pharm. Bull. 40, 1216–1220.
Ishiguro, M., Tanabe, S., Matori, Y., Sakakibara, R., 1992b. Biochemical studies on oral toxicity of ricin. IV. A fate of orally administered ricin in rats. J. Pharmacobiodyn. 15, 147–156.
Jaroszewski, L., 2009. Protein structure prediction based on sequence similarity. Methods Mol. Biol. 569, 129–156.
Jaschke, P.R., Saer, R.G., Noll, S., Beatty, J.T., 2011. Modification of the genome of *Rhodobacter sphaeroides* and construction of synthetic operons. Methods Enzymol. 497, 519–538.
Jones, J.D.G., Dangl, J.L., 2006. The plant immune system. Nature 444, 323–329.
Juberg, D.R., Herman, R.A., Thomas, J., Brooks, K.J., Delaney, B., 2009. Acute and repeated dose (28 day) mouse oral toxicology studies with Cry34Ab1 and Cry35Ab1 Bt proteins used in coleopteran resistant DAS-59122-7 corn. Regul. Toxicol. Pharmacol. 54, 154–163.
Kluh, I., Horn, M., Hýblová, J., Hubert, J., Dolečková-Marešová, L., Voburka, Z., Kudlíková, I., Kocourek, F., Mareš, M., 2005. Inhibitory specificity and insecticidal selectivity of alpha-amylase inhibitor from *Phaseolus vulgaris*. Phytochemistry 66, 31–39.
Kramer, L.C., Choudoir, M.J., Wielgus, S.M., Bhaskar, P.B., Jiang, J., 2009. Correlation between transcript abundance of the *RB* gene and the level of the *RB*-mediated late blight resistance in potato. Mol. Plant Microbe Interact. 22, 447–455.
Kroes, R., Galli, C., Munro, I., Schilter, B., Tran, L.A., Walker, R., Würtzen, G., 2000. Threshold of toxicological concern for chemical substances present in the diet: a practical tool for assessing the need for toxicity testing. Food Chem. Toxicol. 38, 255–312.
Kroes, R., Renwick, A.G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., Schilter, B., Schlatter, J., van Schothorst, F., Vos, J.G., Würtzen, G., 2004. Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food Chem. Toxicol. 42, 65–83.
Kruizinga, A.G., Briggs, D., Crevel, R.W.R., Knulst, A.C., van den Bosch, L.M., Houben, G.F., 2008. Probabilistic risk assessment model for allergens in food: sensitivity analysis of the minimum eliciting dose and food consumption. Food Chem. Toxicol. 46, 1437–1443.
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 Biotechnol. J. 11, 781–784.
Kukuruzinska, M.A., Lennon, K., 1998. Protein N-glycosylation: molecular genetics and functional significance. Crit. Rev. Oral Biol. Med. 9, 415–448.
Lack, G., Chapman, M., Kalsheker, N., King, V., Robinson, C., Venables, K., 2002. Report on the potential allergenicity of genetically modified organsims and their products. Clin. Exp. Allergy 32, 1131–1143.
Ladics, G.S., Cressman, R.F., Herouet-Guicheney, C., Herman, R.A., Privalle, L., Song, P., Ward, J.M., McClain, S., 2011. Bioinformatics and the allergy assessment of agricultural biotechnology products: Industry practices and recommendations. Regul. Toxicol. Pharmacol. 60, 46–53.
Lafont, J., Rouanet, J.M., Gabrion, J., Assouad, J.L., Zambonino Infante, J.L., Besançon, P., 1988. Duodenal toxicity of dietary *Phaseolus vulgaris* lectins in the rat: an integrative assay. Digestion 41, 83–93.
Laible, P.D., Scott, H.N., Henry, L., Hanson, D.K., 2004. Towards higher-throughput membrane protein production for structural genomics initiatives. J. Struct. Funct. Genomics 5, 167–172.
Lardizabal, K.D., Mai, J.T., Wagner, N.W., Wyrick, A., Voelker, T., Hawkins, D.J., 2001. DGAT2 is a new diacylglycerol acyltransferase gene family: purification, cloning, and expression in insect cells of two polypeptides from *Mortierella ramanniana* with diacylglycerol acyltransferase activity. J. Biol. Chem. 276, 38862–38869.
Lee, R.Y., Reiner, D., Dekan, G., Moore, A.E., Higgins, T.J.V., Epstein, M.M., 2013. Genetically modified α-amylase inhibitor peas are not specifically allergenic in mice. PLoS ONE 8, e52972.
Lehrer, S.B., Ayuso, R., Reese, G., 2002. Current understanding of food allergens. Ann. N. Y. Acad. Sci. 964, 69–85.
Lerouge, P., Cabanes-Macheteau, M., Rayon, C., Fischette-Lainé, A.C., Gomord, V., Faye, L., 1998. N-glycoprotein biosynthesis in plants: recent developments and future trends. Plant Mol. Biol. 38, 31–48.
Lin, S.H., Guidotti, G., 2009. Purification of membrane proteins. Methods Enzymol. 463, 619–629.
Liu, L., White, M.J., MacRae, T.H., 1999. Transcription factors and their genes in higher plants: Functional domains, evolution and regulation. Eur. J. Biochem. 262, 247–257.
Luderer, R., Rivas, S., Nürnberger, T., Mattei, B., Van den Hooven, H.W., Van der Hoorn, R.A.L., Romeis, T., Wehrfritz, J.M., Blume, B., Nennstiel, D., Zuidema, D., Vervoort, J., De Lorenzo, G., Jones, J.D.G., De Wit, P.J.G.M., Joosten, M.H.A.J., 2001. No evidence for binding between resistance gene product Cf-9 of tomato and avirulence gene product AVR9 of *Cladosporium fulvum*. Mol. Plant Microbe Interact. 14, 867–876.
Lung, S.C., Weselake, R.J., 2006. Diacylglycerol acyltransferase: a key mediator of plant triacylglycerol synthesis. Lipids 41, 1073–1088.
Macher, B.A., Yen, T.Y., 2007. Proteins at membrane surfaces—a review of approaches. Mol. Biosyst. 3, 705–713.
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 Chem. Toxicol. 50, 3776–3784.
Marsh, D., Horváth, L.I., Swamy, M.J., Mantripragada, S., Kleinschmidt, J.H., 2002. Interaction of membrane-spanning proteins with peripheral and lipid-anchored membrane proteins: perspectives from protein-lipid interactions. Mol. Membr. Biol. 19, 247–255.
Marsh, J.T., Tryfona, T., Powers, S.J., Stephens, E., Dupree, P., Shewry, P.R., Lovegrove, A., 2011. Determination of the *N*-glycosylation patterns of seed proteins: applications to determine the authenticity and substantial equivalence of genetically modified (GM) crops. J. Agric. Food Chem. 59, 8779–8788.
Mathesius, C.A., Barnett Jr., J.F., Cressman, R.F., Ding, J., Carpenter, C., Ladics, G.S., Schmidt, J., Layton, R.J., Zhang, J.X.Q., Appenzeller, L.M., Carlson, G., Ballou, S., Delaney, B., 2009. Safety assessment of a modified acetolactate synthase protein (GM-HRA) used as a selectable marker in genetically modified soybeans. Regul. Toxicol. Pharmacol. 55, 309–320.
Matsoukas, I.G., Massiah, A.J., Thomas, B., 2012. Florigenic and antflorigenic signaling in plants. Plant Cell Physiol. 53, 1827–1842.
McHale, L., Tan, X., Koehl, P., Michelmore, R.W., 2006. Plant NBS-LRR proteins: adaptable guards. Genome Biol. 7, 212.
Mithoe, S.C., Menke, F.L.H., 2011. Phosphoproteomics perspective on plant signal transduction and tyrosine phosphorylation. Phytochemistry 72, 997–1006.
Murmu, J., Plaxton, W.C., 2007. Phospho*enol*pyruvate carboxylase protein kinase from developing castor oil seeds: partial purification, characterization, and reversible control by photosynthate supply. Planta 226, 1299–1310.
Nelson, D.E., Repetti, P.P., Adams, T.R., Creelman, R.A., Wu, J., Warner, D.C., Anstrom, D.C., Bensen, R.J., Castiglioni, P.P., Donnarummo, M.G., Hinchey, B.S., Kumimoto, R.W., Maszle, D.R., Canales, R.D., Krolikowski, K.A., Dotson, S.B., Gutterson, N., Ratcliffe, O.J., Heard, J.E., 2007. Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. Proc. Natl. Acad. Sci. USA 104, 16450–16455.
Oakes, J., Brackenridge, D., Colletti, R., Daley, M., Hawkins, D.J., Xiong, H., Mai, J., Screen, S.E., Val, D., Lardizabal, K., Gruys, K., Deikman, J., 2011. Expression of fungal *diacylglycerol acyltransferase2* genes to increase kernel oil in maize. Plant Physiol. 155, 1146–1157.
Ohlrogge, J., Browse, J., 1995. Lipid biosynthesis. Plant Cell 7, 957–970.
Organisation for Economic Co-operation and Development (OECD), 1995. OECD guideline for the testing of chemicals, Test No. 407: Repeated Dose 28-day Oral Toxicity Study in Rodents (Adopted 27.07.1995). Organisation for Economic Co-operation and Development, Paris.
Organisation for Economic Co-operation and Development (OECD), 2001. OECD guideline for testing of chemicals, Test No. 420: Acute Oral Toxicity – Fixed Dose Procedure (Adopted 17.12.2001). Organisation for Economic Co-operation and Development, Paris.
Padgette, S.R., Re, D.B., Barry, G.F., Eichholtz, D.E., Delannay, X., Fuchs, R.L., Kishore, G.M., Fraley, R.T., 1996. New weed control opportunities: development of soybeans with a Roundup Ready gene. In: Duke, S.O. (Ed.), Herbicide-Resistant Crops: Agricultural, Environmental, Economic, Regulatory, and Technical Aspects. CRC Press, Boca Raton, FL, pp. 53–84.
Parrott, W., Chassy, B., Ligon, J., Meyer, L., Petrick, J., Zhou, J., Herman, R., Delaney, B., Levine, M., 2010. Application of food and feed safety assessment principles to evaluate transgenic approaches to gene modulation in crops. Food Chem. Toxicol. 48, 1773–1790.
Paulick, M.G., Bertozzi, C.R., 2008. The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. Biochemistry 47, 6991–7000.

Pearson, W.R., 2000. Flexible sequence similarity searching with the FASTA3 program package. Methods Mol. Biol. 132, 185–219.
Pearson, W.R., Lipman, D.J., 1988. Improved tools for biological sequence comparison. Proc Natl. Acad. Sci. USA 85, 2444–2448.
Pérez-Rodríguez, P., Riaño-Pachón, D.M., Corrêa, L.G.G., Rensing, S.A., Kersten, B., Mueller-Roeber, B., 2010. PlnTFDB: updated content and new features of the plant transcription factor database. Nucleic Acids Res. 38, D822–D827.
Perilli, S., Moubayidin, L., Sabatini, S., 2010. The molecular basis of cytokinin function. Curr. Opin. Plant Biol. 13, 21–26.
Powilleit, F., Breinig, T., Schmitt, M.J., 2007. Exploiting the yeast L-A viral capsid for the *in vivo* assembly of chimeric VLPs as platform in vaccine development and foreign protein expression. PLoS ONE 2, e415.
Prescott, V.E., Campbell, P.M., Moore, A., Mattes, J., Rothenberg, M.E., Foster, P.S., Higgins, T.J.V., Hogan, S.P., 2005. Transgenic expression of bean α-amylase inhibitor in peas results in altered structure and immunogenicity. J. Agric. Food Chem. 53, 9023–9030.
Privé, G.G., 2007. Detergents for the stabilization and crystallization of membrane proteins. Methods 41, 388–397.
Rayapuram, C., Jensen, M.K., Maiser, F., Shanir, J.V., Hornshøj, H., Rung, J.H., Gregersen, P.L., Schweizer, P., Collinge, D.B., Lyngkjær, M.F., 2012. Regulation of basal resistance by a powdery mildew-induced cysteine-rich receptor-like protein kinase in barley. Mol. Plant Pathol. 13, 135–147.
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 Res. 22, 445–460.
Rayon, C., Lerouge, P., Faye, L., 1998. The protein *N*-glycosylation in plants. J. Exp. Bot. 49, 1463–1472.
Resh, M.D., 1996. Regulation of cellular signalling by fatty acid acylation and prenylation of signal transduction proteins. Cell. Signal. 8, 403–412.
Rice, E.A., Lee, T.C., Rogan, G., Bannon, G.A., 2008. Safety assessment of proteins used in crops developed through agricultural biotechnology: industry perspective. In: Hammond, B.G. (Ed.), Food Safety of Proteins in Agricultural Biotechnology. CRC Press, Boca Raton, FL, pp. 237–257.
Roberts, M.R., 2003. 14-3-3 Proteins find new partners in plant cell signalling. Trends Plant Sci. 8, 218–223.
Rodríguez-Banqueri, A., Kowalczyk, L., Palacín, M., Vázquez-Ibar, J.L., 2012. Assessment of membrane protein expression and stability using a split green fluorescent protein reporter. Anal. Biochem. 423, 7–14.
Roos, C., Zocher, M., Müller, D., Münch, D., Schneider, T., Sahl, H.G., Scholz, F., Wachtveitl, J., Ma, Y., Proverbio, D., Henrich, E., Dötsch, V., Bernhard, F., 2012. Characterization of co-translationally formed nanodisc complexes with small multidrug transporters, proteorhodopsin and with the *E. coli* MraY translocase. Biochim. Biophys. Acta 1818, 3098–3106.
Rossi, M.A., Mancini Filho, J., Lajolo, F.M., 1984. Jejunal ultrastructural changes induced by kidney bean (*Phaseolus vulgaris*) lectins in rats. Br. J. Exp. Pathol. 65, 117–123.
Roy, A., Shukla, A.K., Haase, W., Michel, H., 2008. Employing *Rhodobacter sphaeroides* to functionally express and purify human G protein-coupled receptors. Biol. Chem. 389, 69–78.
Sawada, S., Takeda, Y., Tashiro, M., 2002. Primary structures of α- and β-subunits of α-amylase inhibitors from seeds of three cultivars of *Phaseolus* beans. J. Protein Chem. 21, 9–17.
Saze, H., Ueno, Y., Hisabori, T., Hayashi, H., Izui, K., 2001. Thioredoxin-mediated reductive activation of a protein kinase for the regulatory phosphorylation of C4-form phosphoenolpyruvate carboxylase from maize. Plant Cell Physiol. 42, 1295–1302.
Schlick, T., Collepardo-Guevara, R., Halvorsen, L.A., Jung, S., Xiao, X., 2011. Biomolecular modeling and simulation: a field coming of age. Quart. Rev. Biophys. 44, 191–228.
Schroeder, J.I., Delhaize, E., Frommer, W.F., Guerinot, M.L., Harrison, M.J., Herrera-Estrella, L., Horie, T., Kochian, L.V., Munns, R., Nishizawa, N.K., Tsay, Y.F., Sanders, D., 2013. Using membrane transporters to improve crops for sustainable food production. Nature 497, 60–66.
Shen, B., Allen, W.B., Zheng, P., Li, C., Glassman, K., Ranch, J., Nubel, D., Tarczynski, M.C., 2010. Expression of *ZmLEC1* and *ZmWRI1* increases seed oil production in maize. Plant Physiol. 153, 980–987.
Shen, H.H., Lithgow, T., Martin, L.L., 2013. Reconstitution of membrane proteins into model membranes: seeking better ways to retain protein activities. Int. J. Mol. Sci. 14, 1589–1607.
Shockey, J.M., Gidda, S.K., Chapital, D.C., Kuan, J.C., Dhanoa, P.K., Bland, J.M., Rothstein, S.J., Mullen, R.T., Dyer, J.M., 2006. Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. Plant Cell 18, 2294–2313.
Skach, W.R., 2009. Cellular mechanisms of membrane protein folding. Nat. Struct. Mol. Biol. 16, 606–612.
Spadaro, D., Yun, B.W., Spoel, S.H., Chu, C., Wang, Y.Q., Loake, G.J., 2010. The redox switch: dynamic regulation of protein function by cysteine modifications. Physiol. Plant. 138, 360–371.
Stevenson, S.E., Woods, C.A., Hong, B., Kong, X., Thelen, J.J., Ladics, G.S., 2012. Environmental effects on allergen levels in commercially grown non-genetically modified soybeans: assessing variation across North America. Front. Plant Sci. 3, 196.
Su, M.G., Lee, T.Y., 2013. Incorporating substrate sequence motifs and spatial amino acid composition to identify kinase-specific phosphorylation sites on protein three-dimensional structures. BMC Bioinformatics 14 (Suppl 16), S2.
Suárez-López, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., Coupland, G., 2001. *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. Nature 410, 1116–1120.
Tameling, W.I.L., Takken, F.L.W., 2008. Resistance proteins: scouts of the plant innate immune system. Eur. J. Plant Pathol. 121, 243–255.
Taylor, S.L., Gendel, S.M., Houben, G.F., Julien, E., 2009. The key events dose-response framework: a foundation for examining variability in elicitation thresholds for food allergens. Crit. Rev. Food Sci. Nutr. 49, 729–739.
Taylor, S.L., Moneret-Vautrin, D.A., Crevel, R.W., Sheffield, D., Morisset, M., Dumont, P., Remington, B.C., Baumert, J.L., 2010. Threshold dose for peanut: risk characterization based upon diagnostic oral challenge of a series of 286 peanut-allergic individuals. Food Chem. Toxicol. 48, 814–819.
Turchetto-Zolet, A.C., Maraschin, F.S., de Morais, G.L., Cagliari, A., Andrade, C.M.B., Margis-Pinheiro, M., Margis, R., 2011. Evolutionary view of acyl-CoA diacylglycerol acyltransferase (DGAT), a key enzyme in neutral lipid biosynthesis. BMC Evol. Biol. 11, 263.
Valderrama-Rincon, J.D., Fisher, A.C., Merritt, J.H., Fan, Y.Y., Reading, C.A., Chhiba, K., Heiss, C., Azadi, P., Aebi, M., DeLisa, M.P., 2012. An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. Nat. Chem. Biol. 8, 434–436.
van der Vossen, E., Sikkema, A., te Lintel Hekkert, B., Gros, J., Stevens, P., Muskens, M., Wouters, D., Pereira, A., Stiekema, W., Allefs, S., 2003. An ancient *R* gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. Plant J. 36, 867–882.
van Ooijen, G., Mayr, G., Kasiem, M.M.A., Albrecht, M., Cornelissen, B.J.C., Takken, F.L.W., 2008. Structure–function analysis of the NB-ARC domain of plant disease resistance proteins. J. Exp. Bot. 59, 1383–1397.
van Ooijen, G., van den Burg, H.A., Cornelissen, B.J.C., Takken, F.L.W., 2007. Structure and function of resistance proteins in solanaceous plants. Annu. Rev. Phytopathol. 45, 43–72.
Vleeshouwers, V.G.A.A., Rietman, H., Krenek, P., Champouret, N., Young, C., Oh, S.K., Wang, M., Bouwmeester, K., Vosman, B., Visser, R.G.F., Jacobsen, E., Govers, F., Kamoun, S., Van der Vossen, E.A.G., 2008. Effector genomics accelerates discovery and functional profiling of potato disease resistance and *Phytophthora infestans* avirulence genes. PLoS ONE 3, e2875.
Wang, Y.H., Chollet, R., 1993. Partial purification and characterization of phospho*enol*pyruvate carboxylase protein-serine kinase from illuminated maize leaves. Arch. Biochem. Biophys. 304, 496–502.
Webster, D.E., Thomas, M.C., 2012. Post-translational modification of plant-made foreign proteins; glycosylation and beyond. Biotechnol. Adv. 30, 410–418.
Weinman, M.D., Allan, C.H., Trier, J.S., Hagen, S.J., 1989. Repair of microvilli in the rat small intestine after damage with lectins contained in the red kidney bean. Gastroenterology 97, 1193–1204.
Yamaguchi, H., Funaoka, H., Iwamoto, H., 1992. Structures of sugar chains of the subunits of an α-amylase inhibitor from *Phaseolus vulgaris* white kidney beans. J. Biochem. 111, 388–395.
Yang, J., Ordiz, M.I., Semenyuk, E.G., Kelly, B., Beachy, R.N., 2012. A safe and effective plant gene switch system for tissue-specific induction of gene expression in *Arabidopsis thaliana* and *Brassica juncea*. Transgenic Res. 21, 879–883.
Yang, W.C., Welsh, J.P., Lee, J., Cooke, J.P., Swartz, J., 2011. Solubility partner IF2 Domain I enables high yield synthesis of transducible transcription factors in *Escherichia coli*. Protein Exp. Purif. 80, 145–151.
Yen, C.L.E., Stone, S.J., Koliwad, S., Harris, C., Farese Jr., R.V., 2008. Thematic review series: glycerolipids. DGAT enzymes and triacylglycerol biosynthesis. J. Lipid Res. 49, 2283–2301.
Young, N.M., Thibault, P., Watson, D.C., Chrispeels, M.J., 1999. Post-translational processing of two α-amylase inhibitors and an arcelin from the common bean, *Phaseolus vulgaris*. FEBS Lett. 446, 203–206.
Zaret, K.S., Stevens, K., 1995. Expression of a highly unstable and insoluble transcription factor in *Escherichia coli*: purification and characterization of the *fork head* homolog HNF3α. Protein Exp. Purif. 6, 821–825.
Zheng, P., Allen, W.B., Roesler, K., Williams, M.E., Zhang, S., Li, J., Glassman, K., Ranch, J., Nubel, D., Solawetz, W., Bhattramakki, D., Llaca, V., Deschamps, S., Zhong, G.Y., Tarczynski, M.C., Shen, B., 2008. A phenylalanine in DGAT is a key determinant of oil content and composition in maize. Nat. Genet. 40, 367–372.