InnerPlant, Inc.

Early Food Safety Evaluation for a Green Fluorescent Protein: GFP

Submitting Company

InnerPlant, Inc.
202 Cousteau Place, Suite 150,
Davis, California, 95618

Submitted by:

Difabachew Belay Kondidie, Ph.D.
InnerPlant, Inc.
202 Cousteau Place, Suite 150,
Davis, California, 95618

Contact Information:

Email: belay.kondidie@innerplant.com
Telephone: (402) 805-7366

Table of Contents

List of Figures

List of Tables

Abbreviations and Definitions

BLAST	protein-protein Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
FARRP	Food Allergy Research and Resource Program
GE	Genetically Engineered
GFP	Green Fluorescent Protein
E. coli	*Escherichia coli*
IFB	Innerplant Fungal Biosensor
IgE	Immunoglobulin E
LanFP1	Coding sequence for the LanFP1 Green Fluorescent Protein from lancelet *Branchiostoma floridae*
LanFP1	The Green Fluorescent Protein produced in *Branchiostoma floridae*
NCBI	National Center for Biotechnology Information (United States of America)
PDB	Protein Database Bank
PIR	Protein Information Resource protein sequence database
PRF	Protein Research Foundation protein sequence database
SDS-PAGE	Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis
SGF	Simulated mammalian gastric fluid containing pepsin

1. A Description of the Purpose or Intended Technical Effect of the LanFP1 Protein

InnerPlant is developing a new data stream for agricultural producers that is fueled by the creation of Genetically Engineered (GE) crops that produce an optical fluorescence signal that rapidly and specifically indicates the presence of various biotic and abiotic stresses (optical biosensors). Importantly, InnerPlant has also developed methodology to detect these optical signals in daylight using remote sensing devices that enable detection from tractors, drones, airplanes, and satellites. The combination of biosensors with scalable remote detection capabilities presents an opportunity to provide crop producers with vastly superior information about biological pressures on the crop such as pathogen infection or insect damage as well as abiotic stresses like macro and micro-nutrient deficiencies in the plant. These new data streams will enable producers to reduce fungicide and insecticide usage by targeting only infected areas of the field and will also increase yields by ensuring that pathogens or insect pests are controlled very early in the infection cycle. In addition, nutrient biosensors will enable a step change in precision agriculture unlocking the opportunity to not only reduce over-application of fertilizers but to optimize inputs on a plant-by-plant level.

The basic concept for each of the biosensors InnerPlant is developing is the same, we identify the genetic pathways that respond specifically to a particular stress using transcriptomic and genomic analyses, we then clone the regulatory elements from those endogenous genes and use them to drive the expression of a fluorescent protein that produces an optical signal that can be detected remotely. In the present New Protein Consultation with the U.S. Food and Drug Administration (FDA), InnerPlant has produced GE soybeans that are designed to express a Green Fluorescent Protein (GFP) specifically in response to very early pathogen infection. Hereafter these soybeans are referred to as InnerPlant Fungal Biosensor soybeans or IFB soybeans.

The IFB soybeans were produced by transformation of soybean tissues from non-transgenic cultivar (cv.) Williams 82. The DNA transferred to the soybean genome includes the *lanFP1* gene from lancelet *Branchiostoma floridae* that encodes the LanFP1 Green Fluorescent Protein (Baumann et al., 2008; Bomati et al., 2014). An identified pathogen inducible promoter was fused to the coding sequence of the LanFP1 protein to produce soybeans that emit a fluorescent signal at the onset of pathogen infection. Expression of LanFP1 utilizes the plant's natural disease response pathways, which are activated within hours post-infection. Therefore, IFB soybeans produce LanFP1 specifically at the onset of pathogen infection and this results in rapid production of an optical fluorescence signal that can be detected in the field. This allows early detection of pathogen infection in the soybean crop and enables application of control measures at the onset of infection to mitigate significant damage to the crop. GFP and other fluorescent protein biosensors represent a step change in crop disease management by utilizing the plant's natural disease

response pathways, which are activated within hours post-infection (Westrick et al. 2019; Cabre et al. 2021; Bueno et al. 2022).

2. Name, Description and Function of the LanFP1 Protein

The green fluorescent protein (GFP) was first discovered in the bioluminescent jellyfish *Aequorea victoria* (Shimomura et al., 1962). The subsequent characterization of GFP revealed that its energy-absorbing core, the chromophore, is self-generated via cyclization of a peptide triplet buried in the interior of a protective β-can protein fold (Ormo et al., 1996; Yang et al., 1996). Once oxidized using molecular oxygen, the chromophore shows high stability and absorbance of high-energy light (blue) that is efficiently re-emitted as fluorescence of lower-energy (green) light over a wide range of conditions.

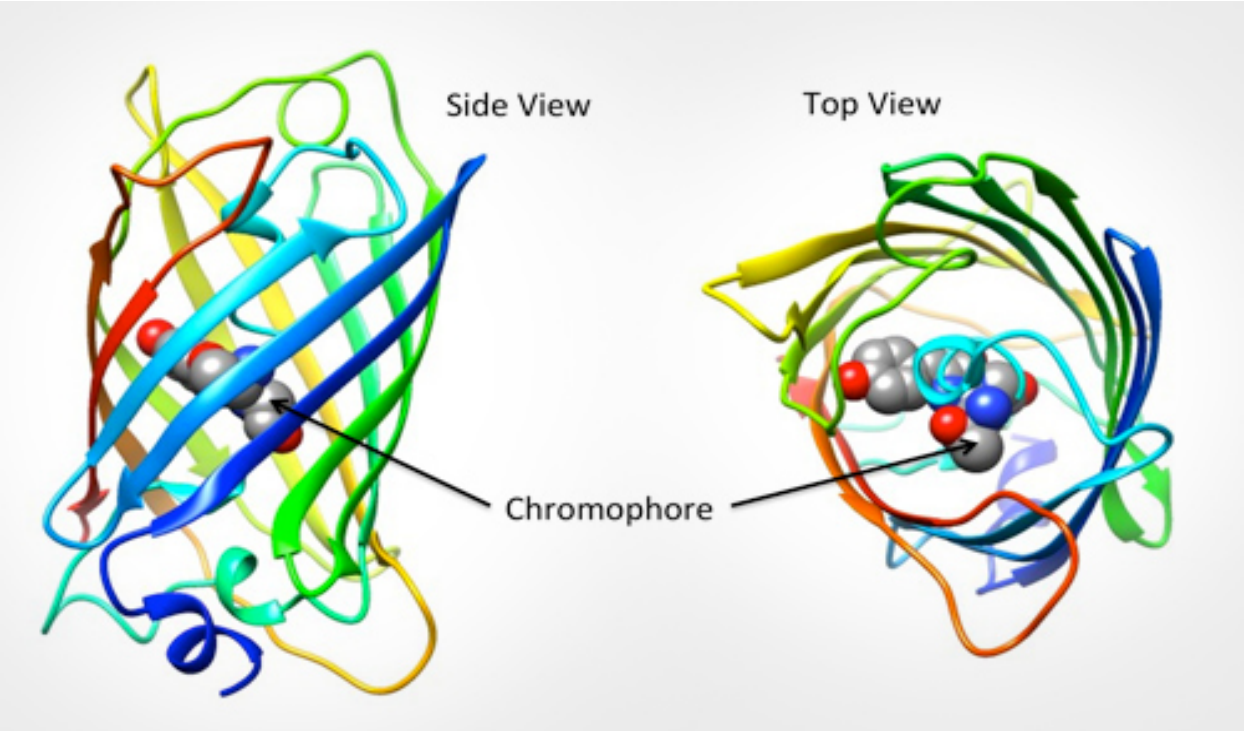
A number of different functions have been proposed for Fluorescent Proteins as summarized by Ong et al. (2011). Because the proteins are colored they might be involved in camouflage of the organism. Also, since most Fluorescent Proteins in corals are located in the same areas as the photosynthetic apparatus it has been proposed that they serve a photoprotective function. Because there are examples of an increase in superoxide dismutase-like activity when some GFPs are exposed to high levels of superoxide it has been suggested that GFPs might provide antioxidant protection. Finally, it has been proposed that Fluorescent Proteins may function as primitive proton pumps and light induced electron-donors.

More recently a family of 16 GFP-like proteins have been identified in the lancelet *Branchiostoma floridae*), the largest set of GFPs known in a single organism (Baumann et al., 2008; Bomati et al., 2009). This extensive family comprises proteins of drastically differing fluorescence intensities and absorbance spectra. Bomati et al. (2009) proposed that some members have light-related functions with a true fluorescence outcome or with only efficient light absorption (e.g., for photoprotection, photoreception) while others have alternative biochemical functions through antioxidant mechanisms (e.g., for cellular defense).

Innerplant introduced the coding sequence for one of these GFPs from *Branchiostoma floridae*., specifically for LanFP1, to produce IFB soybeans. Expression of LanFP1 utilizes the plant's natural disease response pathways, which are activated within hours post-infection. Therefore, IFB soybeans produce LanFP1 specifically at the onset of pathogen infection and this results in rapid production of an optical fluorescence signal that can be detected in the field. This allows early detection of pathogen infection in the soybean crop and enables application of control measures at the onset of infection to mitigate significant damage to the crop. The following describes the mechanism of action of GFPs.

Two independent reports of the x-ray crystal structure of GFP (Ormo *et al*., 1996; Yang *et al*., 1996) revealed that the protein has a unique overall fold comprised of an 11-stranded β-sheet wrapped into a cylindrical β-barrel protein that is 42 amino acids in height and 24 amino acids in diameter (Figure 1). The chromophore is located near the center of the protein, attached to a helical segment of the protein that threads through the center of the β-barrel along its long axis.

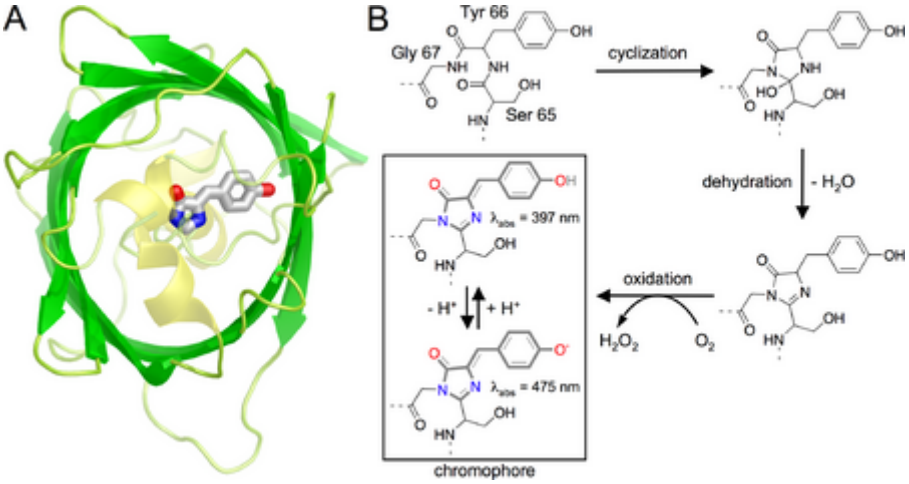
Figure 1. The structure of GFP from the side and top. GFP is a hollow barrel shape with a chromophore in the center (the fluorescent portion). Image reproduced from Protein Database Bank, PDB (2022)



The chromophore is spontaneously formed in GFP within the folded β-barrel protein structure. It has been proposed that formation of the chromophore must necessarily involve at least three key steps: cyclization of the main chain, loss of a molecule of water (dehydration), and oxidation with molecular oxygen (Campbell, 2008). An early, and still generally accepted, proposed mechanism is shown in Figure 2 (Heim et al. 1994). In this mechanism, chromophore formation starts with the nucleophilic glycine 67 amide nitrogen attacking the electrophilic serine 65 carbonyl carbon to form a 5-membered ring in the main chain of the protein. The resulting tetrahedral hemiaminal intermediate undergoes an elimination of water to form a second intermediate. In the final step, the Cα-Cβ bond of tyrosine 66 is oxidized to a double bond with consumption of molecular oxygen and generation of hydrogen peroxide (Zhang et al. 2006). The installation of this double bond simultaneously converts the 5-membered ring into an aromatic system and puts it into conjugation with the aromatic phenol ring of the tyrosine side chain. Chromophore formation is spontaneous

only within the context of the fluorescent protein β-barrel structure where steric constraints force the peptide into a tight turn conformation (Branchini et al. 1998) and the side chains of highly conserved residues, such as glutamate 222 and arginine 96, are positioned to facilitate the reaction.

Figure 2. A. Top view of the GFP structure with barrel shaped protein and central chromophore. **B**. A proposed mechanism for the series of post-translational modifications that converts the serine 65, tyrosine 66, glycine 67 tripeptide sequence into the fluorescent chromophore (Heim et al. 1994). Reproduced from Campbell (2008).



The GFP chromophore exists as an equilibrating mixture of the neutral phenol (absorbance *λmax* = 397 nm) and anionic phenolate (absorbance *λmax* = 475 nm) (Morise et al. 1974; Heim et al. 1994; Patterson et al. 1997). Regardless of whether excitation is at 397 nm or 475 nm, the fluorescence emission occurs from the anionic phenolate species (fluorescence *λmax* = 504 nm) with a quantum yield of 0.79 (Patterson et al. 1997).

3 Identity and Source of Introduced Genetic Material

The coding sequence for the LanFP1 protein (GenBank Accession No. XP_035658893.1) expressed in IFB soybeans was derived from the lancelet *Branchiostoma floridae* (Baumann et al., 2008; Bomati et al., 2009; Bomati et al., 2014). The predicted amino acid sequence of the LanFP1 protein expressed in IFB soybeans is taken from the plasmid used to generate IFB soybeans and is presented in Table 1. The protein consists of 219 amino acids and has a molecular weight of 24,524.73 Da. The amino acid sequence of the LanFP1 protein expressed in IFB soybeans is identical to the sequence of the protein produced in *Branchiostoma floridae* as published by Bomati et al. (2009), except the proline at position 2 of the native sequence is replaced by alanine in the IFB soybeans.

Table 1. Deduced amino acid sequence of the LanFP1 protein produced in IFB soybeans.

IFB Soy	MALPATHDIHLHGSINGHEFDMVGGGKGDPNAGSLVTTAKSTKGALKFSPYLMIPHLGYG
IFB Soy	YYQYLPYPDGPSPFQTSMLEGSGYAVYRVFDFEDGGKLTTEFKYSYEGSHIKADMKLMGS
IFB Soy	GFPDDGPVMTSQIVDQDGCVSKKTYLNNNTIVDSFDWSYNLQNGKRYRARVSSHYIFDKP
IFB Soy	FSADLMKKQPVFVYRKCHVKASKTEVTLDEREKAFYELA

4 Assessment of Allergenicity Potential of the LanFP1 Protein

Large quantities of proteins from diverse sources are consumed daily by humans and are required for a well-balanced and healthy diet. The instance of allergenicity among the tens of thousands of different proteins consumed by humans is rare (Taylor, 1992). There are no definitive methods to assess potential allergenicity of proteins originating from sources not known to produce food allergy. However, there are some recognized procedures that can be used to evaluate a new protein to assess its allergenic potential. Included in these procedures are:

1) Similarity of the new protein to known food allergens with respect to amino acid sequence;
2) Resistance of the protein to digestive degradation;
3) Stability of the protein to heat treatment, and
4) Assessment of the *LanFP1* gene source and history of use and exposure.

These procedures together form the basis for evaluating whether a given protein is likely to be, or become, an allergen (Codex Alimentarius Commission, 2003; Astwood and Fuchs, 1996). Even known food allergens do not always share a group of characteristics and a weight of evidence approach must be used for the protein safety determination**.**

4.1 Amino Acid Sequence Homology of the LanFP1 Protein to Known Protein Allergens

In an assessment of potential protein allergenicity, it is important to establish that the protein does not share potentially immunologically relevant amino acid sequence segments or structure with known allergens. The assessment of potential amino acid sequence homology between the LanFP1 protein and known protein allergens followed the guidelines described in the Codex Alimentarius Commission (2003) report. Two different databases were used to compare the deduced amino acid sequence of the LanFP1 protein to the amino acid sequences of known protein allergens: the Food Allergy Research and Resource Program (FARRP) Allergen Protein Database (version 22.00; allergenonline.com, updated May 25, 2023) and the Compare database (Comprehensive Protein Allergen Resource, version January 25, 2024). First the entire LanFP1 amino acid sequence was compared in the FARRP Allergen Protein database to amino acid sequences of all proteins in this database. Second the LanFP1 sequence was subdivided into overlapping 80-amino acid segments,

and each of these 80 amino acid segments was compared *in silico* to all proteins in both databases. Potential matches between the amino acid sequences of the LanFP1 and proteins in the allergen database were assessed using the FASTA sequence alignment algorithm (Pearson and Lipman, 1988) with default parameters of the database. Any protein showing 35% or greater identity either over the entire sequence or over 80 amino acid overlapping segments to a known allergen would be identified as potentially requiring additional studies. The analyzed LanFP1 protein did not show 35% or greater identity over the whole amino acid sequence of the protein or over 80 amino acid overlapping segments to a potential allergen.

Furthermore, the LanFP1 protein sequence was then submitted to a second search of all possible eight-amino acid subsegments of the query protein against all possible eight-amino acid segments in proteins in the two databases. This eight-amino acid search was based on the concept that eight or more amino acids is a representative minimal size for an IgE-binding epitope (Metcalfe et al., 1996). Bannon and Ogawa (2006) compiled a list of characterized linear IgE-binding epitopes from major allergens and, although one epitope from a wheat ω-5 gliadin was only four amino acids long, the majority of characterized epitopes were indeed eight amino acids or longer. Regions of at least eight consecutive amino acids which are identical between a submitted protein and a known allergen will be identified by this search. The submitted LanFP1 protein sequence did not share a sequence of eight or more consecutive identical amino acids with a potential allergen.

The LanFP1 amino acid sequence did not show 35% or greater identity over the whole amino acid sequence of the protein or over 80 amino acid overlapping segments to a potential allergen. Further, the LanFP1 protein amino acid sequence did not show identity of eight or more consecutive identical amino acids with a potential allergen. Thus, the bioinformatics analyses of the LanFP1 protein expressed in IFB soybeans does not provide any indication of a potential allergenicity concern.

4.2 Lability of the LanFP1 Protein to Pepsin in Simulated Gastric Fluid (SGF)

For a food protein to elicit an allergenic response, the protein must survive the proteolytic environment of the gastrointestinal tract and be absorbed by the intestinal mucosa and initiate an IgE-mediated series of responses. Therefore, evaluating resistance to digestive degradation is one of several criteria used in a weight of evidence approach to assess potential allergenicity of a newly-expressed protein in a genetically-modified plant. The purpose of this study was to demonstrate that the LanFP1 protein has the same susceptibility to digestion in pepsin as other dietary proteins with a history of safe use in foods.

The LanFP1 protein used in this study was prepared from a recombinant *Escherichia coli* over-expression system and was intended for use in a number of studies to confirm the food safety of the protein. The predicted amino acid sequences of the *E. coli*-produced LanFP1 and the LanFP1 protein expressed in IFB soybeans were the same except for the His-Tag residues (LEHHHHHH) at the C-terminal end of the *E. coli*-produced protein (added to facilitate protein purification).

Furthermore, the *E. coli* recombinant LanFP1 protein and the soybean produced protein had the same functional activity. Both sources of the protein produced a bright green fluorescence with blue light excitation, and therefore the two proteins are functionally equivalent. Based on the functional equivalence of the LanFP1 proteins produced in *E. coli* and in soybean as well as on equivalence of the predicted amino acid sequences of the two sources of the LanFP1 protein, the *E.coli*-produced protein is a valid surrogate for the soybean produced protein in the Simulated Gastric Fluid (SGF) studies described in this section of the submission.

The LanFP1 protein was subjected to digestion in SGF. The pepsin digest assay was also run on Bovine Serum Albumin-fraction V (BSA) as a positive control and for comparison to previously published results (Thomas et al., 2004; Astwood and Fuchs, 1996). The lability of the BSA and LanFP1 proteins in SGF was tested by the method described by Thomas et al. (2004). The control and test proteins were subjected to pepsin digest for 0, 0.5, 2, 5, 10 and 20 minutes, and the integrity of the proteins was measured using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Degradation of the BSA control protein and test substance LanFP1 in SGF are presented in Figures 3A and B, respectively. BSA was stable in buffer only (with no pepsin) for 20 minutes, and in buffer with pepsin the full-length protein was rapidly digested within 30 seconds, consistent with previously reported results (Thomas et al., 2004; Astwood and Fuchs, 1996). Some lower molecular protein degradation products were observed across the time course of the assay (Figure 3A). Similarly, the full-length LanFP1 protein was stable in buffer only, but was rapidly degraded in SGF within 30 seconds, and some lower molecular weight protein degradation products were observed across the time course of the assay (Figure 3B). Some of the lower molecular weight staining protein bands were attributed to the pepsin preparation and did not change over the course of the 20-minute assay. Pepsin was stable over the course of the 20-minute assay.

Figures 3A and B. Lability of BSA and LanFP1 to digestion in simulated mammalian gastric Fluid (SGF).

A. With BSA as the test protein.

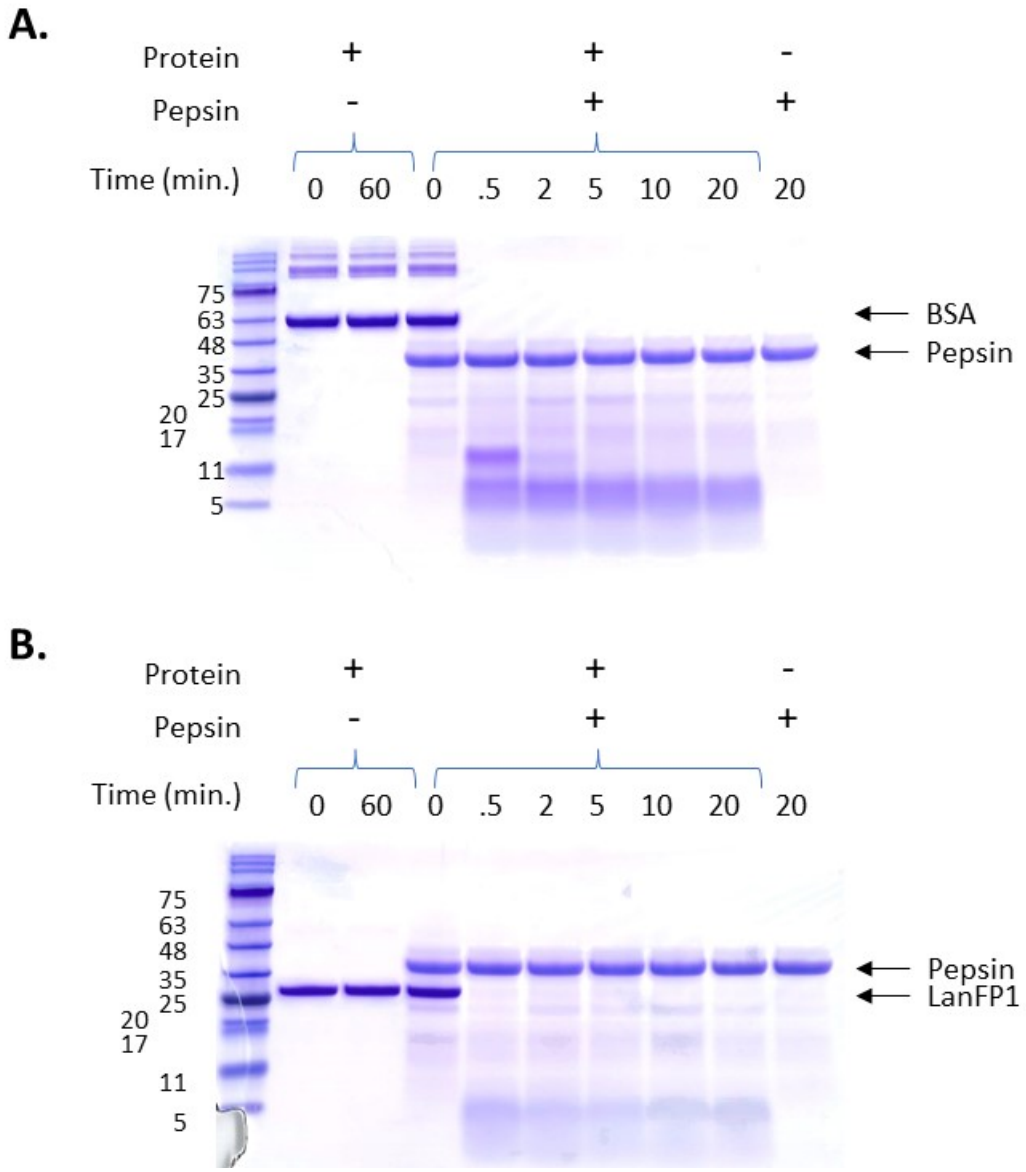
B. With LanFP1 as the test protein.

The first lane in each gel was loaded with the BLUEstain™ 2 Protein ladder (Gold Biotechnology Cat. No. P008-500).

Lanes 2 and 3: Protein in buffer only (no pepsin), incubated from 0 to 60 minutes

Lanes 4-9: Protein incubated in pepsin for 0, 0.5, 2, 5, 10 and 20 minutes

Lane 10: Pepsin only (no BSA or LanFP1 protein), incubated for 20 minutes



N

The *E. coli*-produced LanFP1 protein was rapidly degraded in SGF. This result is consistent with other reports that GFPs are rapidly degraded in SGF (Richards et al., 2003). Therefore, the LanFP1 protein expressed in IFB soybeans is also expected to be rapidly digested the same as conventional dietary proteins. Using a weight of evidence approach for evaluating whether a given protein is likely to be, or become, an allergen (Codex Alimentarius Commission, 2003; Astwood and Fuchs, 1996), one of these criteria is resistance of the protein to pepsin digestion. Results of the current study show that the LanFP1 protein expressed in IFB soybeans is highly digestible under simulated gastric digestion conditions, which is typical of most proteins exposed to the proteases of the mammalian digestive tract. Therefore, based on this criterion for assessment of potential allergenicity, it is unlikely that the LanFP1 protein poses a risk of allergenicity from human consumption.

4.3 Susceptibility of the LanFP1 Protein to Heat Treatment

Most conventional dietary proteins with a history of safe use in food and feed products are denatured by heat treatment during cooking or processing of the raw agricultural commodity to produce food and feed products. For example, cooking or heating is involved in production of a number of soybean processed products, including defatted toasted meal and soy sauce. However, there are examples of some major protein allergens that remain stable to processing to produce different food products (Metcalfe et al., 1996). Therefore, stability to heat treatment is one of the criteria used in a weight of evidence approach to assess potential allergenicity of a protein**.** The purpose of this study was to demonstrate that the LanFP1 protein has the same susceptibility to heat treatment as other dietary proteins with a history of safe use in foods.

The LanFP1 protein used in this study was prepared from a recombinant *Escherichia coli* over-expression system and is the same LanFP1 protein described above in the digestive fate study. The LanFP1 protein is a fluorescent protein with an excitation maximum of 500 nm and an emission maximum at 512 nm (Bomati et al., 2014). This fluorescence is the primary functional activity of LanFP1 and is the basis for the Innerplant detection technology. The impact of heating LanFP1 was monitored by fluorescence imaging of the protein after incubations at 60°, 75°, or 90°C for 10, 30, or 60 minutes. The assay was performed in triplicate. A control sample was held at 4°C. Samples were visualized in a fluorescence imaging station equipped with a Basler ACA3088-57 camera. Fluorescence intensity of the LanFP1 samples treated at 60°C was quantified with ImageJ (Image Processing and Analysis in Java).

Fluorescence of the LanFP1 protein was stable at 4°C over the entire treatment period of 60 minutes (Figure 4A). The fluorescence activity of the LanFP1 protein was stable at 60°C in incubations of 10 and 30 minutes (Figures 4A and B). After a 60-minute incubation at 60°C, approximately 12% of the fluorescence activity was lost (p-value=0.046; two-tailed t-test). Fluorescence from the samples heated to 75° or 90° C were not significantly higher than the buffer only control and were not detectable at the first assay time point (10 minutes). These results

indicate that like most non-thermostable enzymes LanFP1 functional activity is rapidly heat inactivated at temperatures above 60°C. Therefore, functional activity of the LanFP1 protein shows the same sensitivity to heat treatment as other conventional dietary proteins with a history safe consumption in food products.

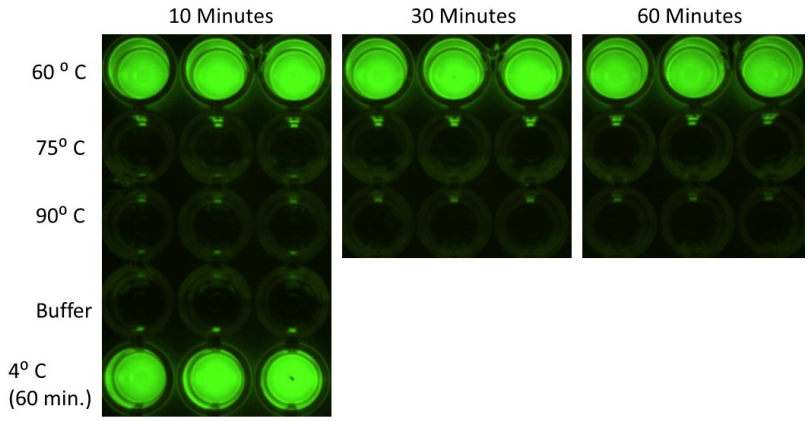
Figure 4. Susceptibility of the *E. coli* produced LanFP1 protein to heat treatments

The fluorescence of the LanFP1 protein was monitored during incubation of a solution of the protein preparation at 4, 60, 75 and 90°C over 60 minutes. The assay was performed in triplicate.

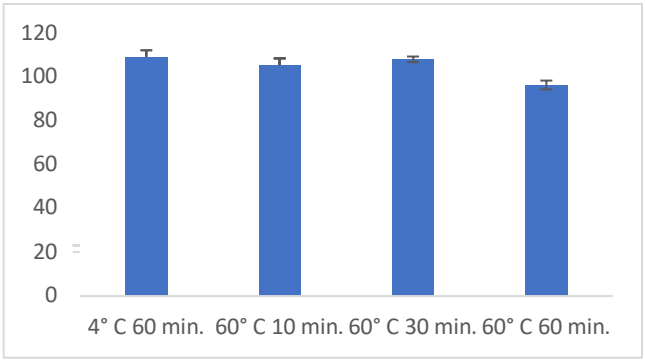
A. Samples were visualized in a fluorescence imaging station equipped with a Basler ACA3088-57 camera (stock no. 107403) fitted with an Edmunds 8.5 mm, f/1.3 Ci series fixed focal length lens (stock no. #86-599) and an Edmunds 520 nm CWL, 25 mm Dia, 36 nm bandwidth, OD 6 fluorescence filter (stock no. #67-030). The excitation light was provided by Cree XEGABL Blue LEDs fitted with a custom filter manufactured by Coherent Labs (center wavelength: 436 nm; bandwidth: 106 nm).

B. Fluorescence intensity of the LanFP1 samples treated at 60°C was quantified with ImageJ (Image Processing and Analysis in Java). Error bars are the standard error.

A.



B.



The functional activity of the LanFP1 protein is rapidly heat inactivated at temperatures above 60°C, the same as conventional dietary proteins. Using a weight of evidence approach for evaluating whether a given protein is likely to be, or become, an allergen (Codex Alimentarius Commission, 2003; Astwood and Fuchs, 1996), one of these criteria is resistance to heat treatment typically during processing of the raw agricultural commodity to food and feed products. Results of the current study show that the functional activity of the LanFP1 protein has the same sensitivity to heat treatment as other conventional dietary proteins with a history safe consumption in food products. Therefore, based on this criterion for assessment of potential allergenicity, it is unlikely that the LanFP1 protein poses a risk of allergenicity from human consumption.

4.4 LanFP1 Gene Source and History of Exposure

The coding sequence for the LanFP1 protein expressed in IFB soybeans was derived from the lancelet *Branchiostoma floridae* (Baumann et al., 2008; Bomati et al., 2009; Bomati et al., 2014). Lancelets are filter feeding, fish-like, marine chordates that resemble anchovies. They are widely distributed in coastal areas of the U.S.A., especially in the Southeast, and have been used as a food source in other countries (Frick and Ruppert, 2001). Lancelets have been harvested and eaten by generations of humans in southeastern China, parts of Taiwan and elsewhere in South East Asia. Frick and Ruppert (2001) reported on the excellent nutritional value of *Branchiostoma floridae* as a food/feed source, but it is not known if *B. floridae* is used for food and feed purposes. Therefore, we were unable to verify a history of safe use of *B. floridae* in food and feed and thereby a history of exposure to the LanFP1 protein in the human and animal diet.

4.5 Conclusions on the Allergenicity Potential of the LanFP1 Protein

Innerplant used a weight of evidence approach to the assessment of the allergenicity potential of the LanFP1 protein (Codex Alimentarius Commission, 2003; Astwood and Fuchs, 1996). This approach recognizes that there is no single characteristic of a protein that defines the allergenicity potential of the protein, but rather there are a number of different characteristics that need to be evaluated together to assess the allergenic potential of the protein.

First, a bioinformatics analysis of potential amino acid sequence homology between the LanFP1 protein and known protein allergens was conducted following the guidelines described in the Codex Alimentarius Commission (2003) report. The LanFP1 amino acid sequence did not show 35% or greater identity over the whole amino acid sequence of the protein or over 80 amino acid overlapping segments to a potential allergen. Further, the LanFP1 protein amino acid sequence did not show identity of eight or more consecutive identical amino acids with a potential allergen. Thus, the bioinformatics analyses of the LanFP1 protein expressed in IFB soybeans show that the LanFP1 protein does not share potentially immunologically relevant amino acid sequence segments or structure with known allergens.

Second, evaluating resistance to digestive degradation is one of several criteria used in a weight of evidence approach to assess potential allergenicity of a newly-expressed protein in a genetically-

modified plant. The LanFP1 protein was subjected to pepsin digestion in a simulated gastric fluid assay (SGF). The LanFP1 protein was rapidly degraded in SGF, which is typical of most safe dietary proteins exposed to the proteases of the mammalian digestive tract. Therefore, based on this criterion for assessment of potential allergenicity, it is unlikely that the LanFP1 protein poses a risk of allergenicity from human consumption.

Third, stability to heat treatment is another criterion used in a weight of evidence approach to assess potential allergenicity of a protein. It was shown that the functional activity of the LanFP1 protein (fluorescence) is rapidly heat inactivated at temperatures above 60°C. Therefore, the LanFP1 protein has the same sensitivity to heat treatment as other conventional dietary proteins with a history of safe consumption in food products. Therefore, based on this criterion for assessment of potential allergenicity, it is unlikely that the LanFP1 protein poses a risk of allergenicity from human consumption.

Even though we were unable to verify a history of safe use of *B. floridae* in food and feed and thereby a history of exposure to the LanFP1 protein in the human and animal diet, the LanFP1 protein lacked sequence homology to known allergens, is susceptible to pepsin degradation and is rapidly inactivated by heat treatment, all characteristics of safe dietary proteins in a weight of evidence approach to an assessment of whether or not a protein may pose a risk of allergenicity.

5.0 Assessment of Toxicity Potential of the LanFP1 Protein

The potential toxicity of the LanFP1 protein was assessed by comparison of the amino acid sequence of the LanFP1 protein to publicly available protein amino acid sequences. Two different databases were utilized for the bioinformatics assessment of potential toxicity. First, a Basic Local Alignment Search Tool (BLAST) search (Altschul et al., 1997) using the LanFP1 amino acid sequence was conducted in the ToxinPred2 database, available at https://webs.iiitd.edu.in/raghava/toxinpred2/ (Sharma et al., 2022) on June 20, 2024. ToxinPred2 is a web-based tool developed for predicting the toxicity of proteins. BLAST version 2.2.20 (Feb 08, 2009) was used with a database containing 16,466 sequences and 5,609,2261 letters, updated on Dec 25, 2021. The search results showed no amino acid sequence homology of the LanFP1 protein to known toxins, and it was also predicted that the sequence is non-toxic.

Second, the GenBank non-redundant peptide sequence database was utilized for bioinformatics assessments of potential toxicity. The GenBank non-redundant peptide sequence data was downloaded June 22, 2024 from the National Center for Biotechnology Information (NCBI) website. This database is comprised of all non-redundant GenBank coding sequence translations, protein sequences from NCBI's Reference Sequence Project, sequences derived from the three-dimensional structure from the Brookhaven Protein Data Bank (PDB) {http://www.wwpdb.org/},

the last major release of the SWISS-PROT protein sequence database (http://www.expasy.org/sprot/), the Protein Information Resource (PIR) protein sequence database (http://pir.georgetown.edu/), and the Protein Research Foundation (PRF) protein sequence database (http://www4.prf.or.jp/en/). Innerplant selected for toxic proteins in the NCBI GenBank database, as described by Negi et al. (2017), and an amino acid sequence similarity BLAST search of the LanFP1 protein sequence was performed against the entries in this NCBI GenBank toxic protein dataset to assess the potential toxicity of the LanFP1 protein. Out of the 1,866,946 entries in the toxic protein dataset, only two proteins showed low levels of amino acid sequence homology to the LanFP1 protein sequence (nucleotidyl transferase AbiEii/AbiGii toxin family protein from *Candidatus Parvarchaeota archaeon*, and ETX/MTX2 family pore-forming toxin from *Bacillus wiedmannii*). However, the amino acid sequence homologies of these two proteins to the LanFP1 sequence were not statistically significant and therefore were not considered biologically meaningful.

Further, the safety of GFP has been demonstrated in peer-reviewed literature. Pure GFP and diets containing transgenic canola expressing GFP were fed to young male rats for 26 days to evaluate the potential toxicity of GFP (Richards et al., 2003). Ingestion of GFP did not affect growth, food intake, relative weight of intestine or other organs, or activities of hepatic enzymes in serum. It was concluded that GFP does not present a risk of toxicity.

6.0 Information on History of Safe Consumption of the LanFP1 Protein in Food

As articulated above in Section 4.4, Innerplant was unable to verify a history of safe use of *B. floridae* in food and feed and thereby a history of exposure to the LanFP1 protein in the human and animal diet. However, as reported in Sections 4 and 5 above, results of bioinformatics analyses as well as susceptibility to pepsin digestion and heat treatment all support the conclusion that the LanFP1 protein is unlikely to pose a risk of allergenicity or toxicity.

7.0 Overall Conclusions

The IFB soybeans were produced by transformation of soybean to introduce the coding sequence of a GFP from the lancelet *Branchiostoma floridae* into the soybean genome. This sequence encodes the LanFP1 protein (Baumann et al., 2008; Bomati et al., 2014). InnerPlant has utilized the understanding of the molecular pathways for plant defense against pathogens to develop IFB soybeans. An identified pathogen inducible promoter was fused to the coding sequence of the LanFP1 protein to produce soybeans that emit a fluorescent signal at the onset of pathogen infection. Expression of LanFP1 utilizes the plant's natural disease response pathways, which are

activated within hours post-infection. Therefore, IFB soybeans produce LanFP1 specifically at the onset of pathogen infection and this results in rapid production of an optical fluorescence signal that can be detected in the field. This allows early detection of pathogen infection in the soybean crop and enables application of control measures at the onset of infection to mitigate significant damage to the crop.

The purpose of the New Protein Consultation submission to the FDA is to demonstrate the food and feed safety of the LanFP1 protein expressed in IFB soybeans. Innerplant followed FDA's guidance for the early food safety evaluation of new proteins in new plant varieties (FDA, 2006) to evaluate the LanFP1 protein for its allergenicity and toxicity potential. Innerplant used a weight of evidence approach to the assessment of the allergenicity potential of the LanFP1 protein (Codex Alimentarius Commission, 2003; Astwood and Fuchs, 1996). This approach recognizes that there is no single characteristic of a protein that defines the allergenicity potential of the protein, rather there are a number of different characteristics that need to be evaluated together to assess the allergenic potential of the protein. The criteria assessed were as follows:

- A bioinformatic analysis of the amino acid sequence of the LanFP1 protein compared to the sequence of putative or known allergens;
- Lability of the LanFP1 protein to pepsin in an SGF assay;
- Susceptibility of the LanFP1 protein to heat treatment, and
- An assessment of the *LanFP1* gene source and history of use or exposure to the LanFP1 protein.

A bioinformatics analysis of potential amino acid sequence homology between the LanFP1 protein and known protein allergens was conducted following the guidelines described in the Codex Alimentarius Commission (2003) report. The LanFP1 amino acid sequence did not show 35% or greater identity over the whole amino acid sequence of the protein or over 80 amino acid overlapping segments to a potential allergen. Further, the LanFP1 protein amino acid sequence did not show identity of eight or more consecutive identical amino acids with a potential allergen. Thus, the bioinformatics analyses of the LanFP1 protein expressed in IFB soybeans shows that the LanFP1 protein does not share potentially immunologically relevant amino acid sequence segments or structure with known allergens.

The LanFP1 protein was rapidly degraded in an SGF assay (pepsin), and the functional activity of LanFP1 (fluorescence) was inactivated when the protein was treated at temperatures above 60°C. Therefore, the LanFP1 protein has the same sensitivity to pepsin digestion and heat treatment as other conventional dietary proteins with a history of safe consumption in food products.

Even though we were unable to verify a history of safe use of *B. floridae* in food and feed and thereby a history of exposure to the LanFP1 protein in the human and animal diet, the LanFP1 protein lacked sequence homology to known allergens, is susceptible to pepsin degradation and is

rapidly inactivated by heat treatment, all characteristics of safe dietary proteins in a weight of evidence approach to an assessment of whether or not a protein may pose a risk of allergenicity. Therefore, it is unlikely that the LanFP1 protein poses a risk of allergenicity from human consumption.

Bioinformatic analyses showed no biologically relevant amino acid sequence similarities between the LanFP1 protein and known protein toxins. Additionally, pure GFP and diets containing transgenic canola expressing GFP were fed to young male rats for 26 days, and no adverse effects were observed (Richards et al., 2003). It was concluded that GFP does not present a risk of toxicity. Collectively, these data support the conclusion that the LanFP1 protein is unlikely to have toxic or adverse effects in mammals.

In summary, results of the studies presented in Innerplants' New Protein Consultation submission to the FDA show that the LanFP1 protein expressed in IFB soybeans does not possess any attributes of known food allergens and is not toxic to mammals.

8.0 Literature Cited

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research, 25(17):3389-3402.

Astwood, J. D., and Fuchs, R. L. (1996) Allergenicity of foods derived from transgenic plants. Monogr. Allergy. 32:105-120.

Bannon, G. A. and Ogawa, T. (2006) Evaluation of available IgE-binding epitope data and its utility in bioinformatics. Mol. Nutr. Food Res. 50:638-644.

Baumann, D., Cook, M., Ma, L., Mushegian, A., Sanders, E., Schwartz, J., and Yu, C. R. (2008) A family of GFP-like proteins with different spectral properties in lancelet *Branchiostoma floridae*. Biol Direct. 3:28.

Bomati, E. K., Manning, G., and Deheyn, D. D. (2009) Amphioxus encodes the largest known family of green fluorescent proteins, which have diversified into distinct functional classes. BMC Evolutionary Biology 9:77.

Bomati, E. K., Haley, J. E., Noel, J. P., and Deheyn, D. D. (2014) Spectral and structural comparison between bright and dim green fluorescent proteins in Amphioxus. Sci Rep. 27:5469.

Branchini, B. R., Nemser, A. R., and Zimmer, M. (1998) A computational analysis of the unique protein-induced tight turn that results in posttranslational chromophore formation in green fluorescent protein. J. Am. Chem. Soc. 120:1-6.

Bueno, T. V., Fontes, P. P., Abe, V. Y., Saito, A. U., Senra, R. L., Oliveira, L. S., Dos Santos, A. B., Capote Ferreira, E. G., Darben, L. M., de Oliveira, A. B., Abdelnoor, R. V., Whitham, S. A., Fietto, L. G., and Marcelino Guimarães, F. C. (2022) A *Phakopsora pachyrhizi* effector suppresses PAMP-triggered immunity and interacts with a soybean glucan endo-1,3- β -glucosidase to promote virulence. Molecular plant-microbe interactions MPMI, 10.1094/MPMI-12-21-0301-R. Advance online publication. https://doi.org/10.1094/MPMI-12-21-0301-R.

Cabre, L., Peyrard, S., Sirven, C., Gilles, L., Pelissier, B., Ducerf, S., and Poussereau, N. (2021) Identification and characterization of a new soybean promoter induced by *Phakopsora pachyrhizi*, the causal agent of Asian soybean rust. BMC Biotechnol. 21(1):27.

Campbell, R. E. (2008) Fluorescent proteins. Scholarpedia, 3(7):5410.

Codex Alimentarius Commission. (2003) Alinorm 03/34: Joint FAO/WHO Food Standard Programme, Codex Alimentarius Commission, Twenty-Fifth Session, Rome, Italy, 30 June-5 July, 2003. Appendix III, Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants, and Appendix IV, Annex on the assessment of possible allergenicity, pp. 47-60.

FDA. (2006) Guidance for Industry: Recommendations for the early food safety evaluation of new non-pesticidal proteins produced by new plant varieties intended for food use. 71 FR 35688.

Frick, J. E., and Ruppert, E E. (2001) Preliminary nutritional analysis of lancelets, a promising seafood with aquacultural potential. Journal of Aquatic Food Product Technology 10:(1):63-75.

Heim, R., Prasher, D. C., and Tsien, R. Y. (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc. Natl. Acad. Sci. U.S.A. 91:12501-12504.

Metcalfe, D. D., Astwood, J. D., Townsend, R., Sampson, H. A., Taylor S. L., and Fuchs, R. L. (1996) Assessment of the allergenic potential of foods derived from genetically engineered crop plants. Critical Reviews in Food Science and Nutrition 36:S165-S186.

Morise, H., Shimomura, O., Johnson, F. H., and Winant, J. (1974) Intermolecular energy transfer in the bioluminescent system of Aequorea. Biochemistry 13:2656-2662.

Negi, S. S., Schein, C. H., Ladics, G. S., Mirsky, H., Chang, P., Rascle, J-B, Kough, J., Sterck, L., Papineni, S., Jez, J. M., Mouriès, L. P., and Braun, W. (2017) Functional classification of protein toxins as a basis for bioinformatic screening. Scientific Reports 7:13940.

Ong, W. J.-H., Alvarez, S., Leroux, I. E., Shahid, R. S., Samma, A. A., Peshkepija, P., Morgan, A. L., Mulcahy, S., and Zimmer, M. (2011) Function and structure of GFP-like proteins in the protein data bank. Mol. BioSyst. 7:984–992.

Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) Crystal structure of the *Aequorea victoria* green fluorescent protein. Science 273:1392-1395.

Patterson, G. H., Knobel, S. M., Sharif, W. D., Kain, S. R., and Piston, D. W. (1997) Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. Biophys. J. 73:2782-2790.

Pearson, W. R., and Lipman, D. J. (1988) Improved tools for biological sequence comparison. Proceedings of the National Academy of Sciences 85(8):2444-2448.

Protein Database Bank (PDB) (2022) Molecular models: Exploring the structure of fluorescent proteins. PDB 101

Richards, H. A., Han, C-T., Hopkins, R. G., Failla, M. L., Ward, W. W., and Stewart, Jr. C. N. (2003) Safety assessment of recombinant green fluorescent protein orally administered to weaned rats. J. Nutr. 133:1909–1912.

Sharma, N., Naorem, L. D., Jain, S., and Raghava, G. (2022) ToxinPred2: an improved method for predicting toxicity of proteins. Briefings in bioinformatics, bbac174. Advance online publication. https://doi.org/10.1093/bib/bbac174.

Shimomura, O., Johnson, F. H., and Saiga, Y. (1962) Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. J. Cell Comp. Physiol. 59:223-239.

Taylor, S. L. (1992) Chemistry and detection of food allergens. Food Technol. 46:146-152.

Thomas, K., Aalbers, M., Bannon, G. A., Bartels, M.; Dearman, R. J., Esdaile, D. J., Fu, T. J., Glatt, C. M., Hadfield, N., Hatzos, C., Hefle, S. L., Heylings, J. R., Goodman, R. E., Henry, B., Herouet, C., Holsapple, M., Ladics, G. S., Landry, T. D., MacIntosh, S. C., Rice, E. A., Privalle, L. S., Steiner, H. Y., Teshima, R., van Ree, R., Woolhiser, M., and Zawodny, J. (2004) A multi-laboratory evaluation of a common *in vitro* pepsin digestion assay protocol used in assessing the safety of novel proteins. Regulatory Toxicology and Pharmacology 39:87-98.

Westrick, N. M., Ranjan, A., Jain, S., Grau, C. R., Smith, D. L., and Kabbage, M. (2019) Gene regulation of *Sclerotinia sclerotiorum* during infection of *Glycine max*: on the road to pathogenesis. BMC Genomics 20:157.

Yang, F., Moss, L. G., and Phillips, G. N. J. (1996) The molecular structure of green fluorescent protein. Nat. Biotechnol. 14:1246-1251.

Zhang, L., Patel, H. N., Lappe, J. W., and Wachter, R. M. (2006) Reaction progress of chromophore biogenesis in green fluorescent protein. J. Am. Chem. Soc. 128:4766-4772.

InnerPlant, Inc.
202 Cousteau Place, Suite 150,
Davis, California, 95618

Attn: Dr. Matthew L. Fabian
Office of Food Additive Safety
Center for Food Safety and Applied Nutrition
U.S. Food and Drug Administration

Subject: InnerPlant responses to letter dated October 30, 2024, regarding our New Protein Consultation (NPC) submission 000020

November 1, 2024

Dear Dr. Fabian,
Thank you for your letter of October 30, 2024, with some technical questions regarding the Innerplant NPC submission 000020 for the safety of the LanFP1 protein. Responses to these questions are shown below. The question from the U.S FDA is italicized and is followed by the InnerPlant response.

Question 1*: In InnerPlant's submission, you briefly describe the digestion assay conducted to demonstrate that the LanFP1 protein has the same susceptibility to digestion in pepsin as other dietary proteins with a history of safe use. You cite the method by Thomas et al., (2004) but do not explicitly state the ratio of pepsin to protein in your assay. The ratio of pepsin to target protein is a key parameter relevant to evaluation of assay results. Please confirm the ratio of pepsin to protein used in the LanFP1 in vitro digestibility assay described in Section 4.2, pages 10-13.*

In the pepsin digest method described by Thomas et al. (2004), the authors used a ratio of 10 units of pepsin activity to one µg of test protein. Innerplant used an equivalent ratio in our pepsin digest study of the LanFP1 protein described in the NPC submission. Below is an excerpt from the Materials and Methods of the Innerplant Regulatory Science report of this study that provides details on the concentrations of the pepsin and LanFP1 proteins used in the pepsin digest study. As you will see, there were 1974 units of pepsin activity and 200 µg of LanFP1 protein in each digest, so the ratio of pepsin activity to LanFP1 protein is 9.87 to 1.

Simulated mammalian gastric fluid reactions.

The lability of the recombinant LanFP1 protein in simulated gastric fluid (SGF) was tested by the method described by Thomas et al. (2004). The assays consisted of approximately 10 U of pepsin/µg of a LanFP1 protein with a C-terminal 6x-His tag or the BSA control protein.

Protein digest reagents

Pepsin was purchased from MilliporeSigma (Catalog # P6887; Lot # 0000313133). The activity of the lot is 3,200 U/mg (assay performed by MilliporeSigma). Immediately prior to use, pepsin was dissolved in water to a final concentration of 10 mg/mL or 32,000 U/mL.

Buffer preparation

Buffer A (control buffer without pepsin):

Component	Stock Concentration	Units	Volume (µL)	Final Concentration
Water	-	-	1929	-
NaCl	2500	mM	28	35
HCl	4	N	43	0.086

Buffer B (SGF):

Component	Stock Concentration	Units	Volume (µL)	Final Concentration
Water	-	-	1764.5	-
NaCl	2500	mM	28	35
HCl	4	N	43	0.086
Pepsin	32000	U/mL	164.5	2631.6

Table 1. The complete list of samples used in this study.

Index	Test Protein	Pepsin	Protein (µL)	Buffer	Time (minutes)*
B1	BSA	-	2.7	A	0
B2	BSA	-	2.7	A	20
B3	BSA	+	2.7	B	0
B4	BSA	+	20	B	0.5
B5	BSA	+	20	B	2
B6	BSA	+	20	B	5
B7	BSA	+	20	B	10
B8	BSA	+	20	B	20
B9	-	+	-	B	20
G1	LanFP1	-	2.7	A	0
G2	LanFP1	-	2.7	A	20
G3	LanFP1	+	2.7	B	0
G4	LanFP1	+	20	B	0.5
G5	LanFP1	+	20	B	2
G6	LanFP1	+	20	B	5
G7	LanFP1	+	20	B	10
G8	LanFP1	+	20	B	20
G9	-	+	-	B	20

Testing of LanFP1 and BSA for lability in SGF.

Control assays consisted of 2.7 µl of the test or control protein (10 mg/mL) in 97 µl of buffer A without pepsin and incubated at 37° C for 0 or 20 minutes (index B1 and B2, and G1 and G2 from table above). Another control reaction consisted of the SGF buffer that had been quenched with 200 mM carbonate buffer pH 11.0 prior to the addition of either the BSA or LanFP1 proteins (indices B3 and G3 respectively). At the indicated time, the control reaction mixture was mixed with 35 µL of 200 mM carbonate buffer pH 11. Then 45 µL of 4X LDS loading buffer (Millipore, Cat No. MPSB-10 mL) was added, mixed and the samples were heated at 75° C for 10 minutes.

For the SGF assays, 20 µl of the test or control protein (10 mg/mL) was added to 750 µL buffer B with pepsin that had been pre-warmed to 37° C. Treatment indices B9 and G9 consisted of the SGF buffer B with neither BSA or LanFP1 added, respectively, and incubated for 20 minutes. At the indicated time a 100 µL aliquot of the reaction was removed and mixed with 35 µL of 200 mM carbonate buffer pH 11. Then 45 µL of 4X LDS loading buffer (Millipore, Cat No. MPSB-10 mL) was added, mixed and the samples were heated to 75° C for 10 minutes. Treatment indices B9 and G9 consisted of the SGF buffer B with neither BSA or LanFP1 added, respectively, and incubated for 20 minutes.

Question 2*: In InnerPlant's submission, you state that the safety of GFP has been demonstrated in peer-review literature and you cite Richards et al., (2003), which reports the safety assessment of a purified GFP and of recombinant GFP expressed in canola. You do not explain your basis for concluding that results for the particular GFPs tested in the Richards et al., (2003) study are relevant to your safety assessment of LanFP1. When referencing studies conducted on a substance that is not the subject of your safety assessment, it is important to explain why results from studies of the former are relevant to the latter. Please clarify your rationale for citing the Richards et al., 2003 study on page 17 and elsewhere. How is the tested GFP as described in this study similar to the subject LanFP1 protein?*

Since LanFP1 is a member of the green fluorescent protein (GFP) family that share conserved amino acid sequences and have similar fluorescence modes of action (Baumann et al., 2008), the intent of citing Richards et al. (2003) was to present an example of the safety of a GFP protein but was not intended to imply that the safety of this GFP was equally applicable to the LanFP1 protein.

Baumann, D., Cook, M., Ma, L., Mushegian, A., Sanders, E., Schwartz, J., and Yu, C. R. (2008) A family of GFP-like proteins with different spectral properties in lancelet *Branchiostoma floridae*. Biol Direct. 3:28.

In order to make this clear in the text of NPC 000020, we have amended the language as follows:

Page 17. LanFP1 is a member of the green fluorescent protein (GFP) family that share conserved amino acid sequences and have similar fluorescence modes of action (Baumann et al., 2008). As an example of the safety of a GFP, the absence of toxicity associated with mGFP5ER has been demonstrated in peer-reviewed literature. Pure mGFP5ER and diets containing transgenic canola expressing mGFP5ER were fed to young male rats for 26 days to evaluate the potential toxicity of mGFP5ER (Richards et al., 2003). Ingestion of mGFP5ER did not affect growth, food intake, relative weight of intestine or other organs, or activities of hepatic enzymes in serum. It was concluded that mGFP5ER does not present a risk of toxicity.

Page 19. Bioinformatic analyses showed no biologically relevant amino acid sequence similarities between the LanFP1 protein and known protein toxins. Additionally, the absence of toxicity associated with a GFP has been demonstrated. Pure mGFP5ER and diets containing transgenic canola expressing mGFP5ER were fed to young male rats for 26 days, and no adverse effects were observed (Richards et al., 2003). It was concluded that mGFP5ER does not present a risk of toxicity. Therefore, the bioinformatics analyses and the example of absence of toxicity associated with a GFP support the conclusion that the LanFP1 protein is unlikely to have toxic or adverse effects in mammals.

The revised version of NPC 000020 is included with the responses to your questions, and the revised submission is differentiated from the original by today's date in the file name.

Thank you again for your review of the InnerPlant NPC 000020 submission and please let us know if our responses to your questions are to your satisfaction and if you have additional questions.

Yours Sincerely,

Belay Kondidie,
Stewardship and Regulatory Manager,
InnerPlant, Inc.