InnerPlant, Inc.

Regulatory Status Review Request for the Determination of Nonregulated Status for Soybean Constitutively Expressing a Fluorescent Marker Protein

The purpose of this Regulatory Status Review (RSR) is to request a determination that the article should not be regulated under 7 CFR Part 340

Submitting Company

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

Submitted by:

Randall Shultz, Ph.D.
Vice President of Research
InnerPlant, Inc.
202 Cousteau Place, Suite 150,
Davis, California, 95618

Contact Information:

Email: randy.shultz@innerplant.com
Telephone: (510) 543-1818

Table of Contents

List of Figures

List of Tables

Abbreviations and Definitions

[	]	**CBI-deleted**
APHIS	Animal and Plant Health Inspection Service	
CaMV	Cauliflower Mosaic Virus	
CFIA-PBO	Canadian Food Inspection Agency-Plant Biosafety Office	
[	]	**CBI-deleted**
cv.	Cultivar	
[	]	**CBI-deleted**
FOIA	Freedom of Information Act	
GE	Genetically Engineered	
GFP	Green Fluorescent Protein	
HSP	Heat Shock Protein	
ICM	InnerPlant Constitutive Marker	
[	]	**CBI-deleted**
OECD	Organization for Economic Cooperation and Development	
[	]	**CBI-deleted**
RSR	Regulatory Status Review	
[	]	**CBI-deleted**
USDA	United States Department of Agriculture	
UTR	Untranslated Region	

1 Confidential Business Information (CBI) Statement

This RSR request does contain CBI.

2 CBI Justification

The Freedom of Information Act (FOIA) exempts federal agencies from releasing information that is trade secret and commercial or financial information that is privileged or confidential (5 U.S.C. 552(b)(4)). InnerPlant considers certain information in this application as trade secret. Disclosure of this information would cause substantial competitive harm to InnerPlant by allowing other companies to unfairly compete with InnerPlant. InnerPlant must keep its research confidential: what it is doing and how it is doing it. Disclosure of this information would enable competitors to duplicate InnerPlant research and development without incurring the investment of time and money expended by the company. Moreover, InnerPlant must protect its intellectual property. InnerPlant must keep research information strictly confidential because in some cases, patent applications have not been filed or patents are pending and have not been published.

Specifically, InnerPlant designates the following as Confidential Business Information:

Genetic Elements Identity. InnerPlants biotechnology traits consist of vectors transferred into plants, which comprise genes for the expression of traits and regulatory sequences such as promoters, enhancers, signaling peptides and terminators. Disclosure of this information may also reveal the origin of these genes and genetic elements and the specific modifications the company made in assembling the DNA constructs and enhance their usefulness. It is in InnerPlants commercial interest that these trade secrets not be publicly disclosed.

Transformation Methodology. InnerPlant has developed a novel method for transformation and selection of transformed plants. The key components of this highly efficient transformation method are the selectable marker and specific genetic elements associated with expression of the selectable marker in soybean tissues, as well as the transformation method itself. Disclosure of this information would enable competitors to duplicate InnerPlants research and development without the investment InnerPlant has made in developing this method.

3 Product Description and Rationale

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 indicate 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 the industry with vastly superior information about biological pressures such as pathogen infection or insect damage as well as abiotic stresses such as macro and micro-nutrient deficiencies in the soil. These new data streams will enable producers to reduce pesticide 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. This approach does not alter the plant's endogenous metabolic pathways, agronomic characteristics or interactions with the environment, but simply adds the new fluorescence gene in combination with the copy of the regulatory elements. Different biosensor designs are built in the same way using a different set of promoter and regulatory elements that respond to the desired stress condition, and in some cases, we use a different fluorescent protein with unique optical properties (excitation and emission wavelenghts) to enable multiplexing of biosensor signals. An example of one of InnerPlants optical biosensor plants is soybean expressing a fluorescent marker protein in response to pathogen infection of the plant (Regulatory Status Review submission number RSR: 22-235-01rsr).

In the present application InnerPlant is requesting a Regulatory Status Review (RSR) of GE soybeans that are designed to constitutively express a green fluorescent protein (GFP). Although not intended to be a stand-alone commercial product, these plants with constant fluorescent protein expression are an important ancillary to InnerPlants optical biosensor plants because they serve as a critical tool to enable us to calibrate, refine and improve the design of our detection equipment in the field. Hereafter these soybeans are referred to as InnerPlant Constitutive Marker soybeans or ICM soybeans.

The ICM soybeans are being produced by [ ] transformation CBI-deleted
of soybean tissues from non-transgenic cultivar (cv.) Williams 82 using [ ]. CBI-deleted
The DNA transferred from the [ ] contains two gene CBI-deleted
constructs. The first gene is *lanFP1* encoding the Green Fluorescent Protein (GFP). ICM soybeans produce GFP constitutively and this results in production of an optical fluorescence signal that can be detected in the field. The strong, constitutive expression of the GFP protein enables the development, refinement, and calibration of our detection equipment.

The second gene transferred to the soybean genome encodes [CBI-deleted
]. Expression of this protein in plant cells [CBI-deleted CBI-deleted
] and serves as a selectable marker for CBI-deleted
plant transformation. The mechanism of action and safety of the GFP and [ ] proteins are CBI-deleted
reviewed in this RSR application.

4 Description of Comparator Plant

The biology of soybean described herein is based upon the consensus document for *Glycine max* (L.) Merr. prepared by the Organization for Economic Co-operation and Development (OECD, 2000), as well as a biology document published by CFIA-PBO (CFIA, 1996), and recent literature on the topic.

A The Taxonomy and Genetics of Soybean

The cultivated soybean, *Glycine max* (L.) Merr., a diploidized tetraploid (2n=40) that belongs to the family Fabaceae and is further classified taxonomically as follows:

Kingdom	Plantae -- Plants
Subkingdom	Tracheobionta -- vascular plants
Division	Magnoliophyta -- angiosperms, flowering plants
Class	Magnoliopsida -- dicots
Subclass	Rosidae
Order	Fabales
Family	Fabaceae
Genus	Glycine Willd. -- soybean
Species	Glycine max (L.) Merr. -- soybean

The above taxonomic information for soybean was obtained from the Integrated Taxonomic Information System (http://www.itis.gov/) and soybean is assigned the taxonomic serial number 26716.

The genus *Glycine* Willd. contains two subgenera, *Glycine* and *Soja* (Moench) F.J. Herm. The subgenus *Glycine* comprises 22 wild perennial species that are indigenous to Australia, islands in the west, central and southern Pacific Ocean, China, Russia, Japan, Indonesia, Korea, Papua New Guinea, the Philippines, and Taiwan (Hymowitz, 2004) The cultivated soybean, *G. max* (L.) Merr. and its wild annual relatives from Asia, *G. soja* Sieb. and Zucc. are classified in the subgenus *Soja*. *Glycine soja* is an annual that grows in the wild in fields, hedgerows, roadsides, and riverbanks in many countries of East Asia.

In addition to *G. max* and *G. soja*, the subgenus *Soja* also contains a form known as *G. gracilis*. This semi-cultivated or weedy plant is found only in Northeast China and is intermediate in morphology between *G. max* and *G. soja*. *G. gracilis* is a variant of *G. max* (Hermann, 1962; Wang, 1976; Shoemaker *et al*., 1986). The three species of the subgenus *Soja* are capable of cross-pollination and the hybrid seed that is produced can germinate normally and produce plants with

fertile pollen and seed (Singh and Hymowitz, 1989). The wild, weedy relatives of *G. max, G. soja* and *G. gracilis* are indigenous to Asia and do not occur in the U.S. (USDA-APHIS, 2008). Therefore, there is no potential for outcrossing of *G. max* to weedy relatives in the U.S.

B Reproductive Biology and Hybridization with Cultivated Soybean

Soybean is a self-pollinating species that is propagated by seed (OECD, 2000). Due to the strong propensity for self-fertilization, the frequency of soybean cross-pollination is very low. For example, plants grown in close proximity to each other (15 cm) were found to have average outcrossing rates of 1.8%, while plants separated by distances of 0.9 m and 5.4 m had outcrossing rates of 0.41 and 0.03%, respectively (Ray *et al*., 2003). Soybeans are generally not a preferred plant for insect pollinators and insect activity has been found not to increase the outcrossing rate (Erickson, 1975; Erickson, 1984). The regulations governing the production of certified Foundation soybean seed are consistent with the low outcrossing rate recognized for soybean. These regulations place no restriction on the separation distance between different cultivars in the field provided that it is sufficient to prevent mechanical mixing during harvest (USDA-APHIS, 2008).

C Weediness Potential of Cultivated Soybean

Soybean plants are not weedy and are not found outside of cultivation. Soybeans are annuals that reproduce solely from seeds. Cultivated soybean rarely displays any dormancy characteristics (a desirable trait that is selected for in commercial varieties) (TeKrony *et al*., 1987) and are sensitive to cold temperatures (Raper and Kramer, 1987). Therefore, their potential to survive in the U.S. from one growing season to the next is very low. Soybean seeds normally germinate quickly under the appropriate environmental conditions that include adequate moisture and moderate temperatures and could potentially grow as a volunteer. However, any volunteers that grow after harvest would be destroyed by the low and freezing temperatures encountered during the following winter. In the event that volunteers were to become established, they would not compete well with succeeding crops and they could be controlled by either mechanical or chemical means (OECD, 2000). The low weediness potential for soybeans is reflected in the fact that soybeans are not listed on the USDA Federal Noxious Weed List (USDA, 2006).

D Characteristics of the Recipient Soybean Cultivar

The recipient soybean cultivar Williams 82 (PI518671) was genetically engineered to express the fluorescent protein. Cultivar Williams 82 is a Phytophthora-resistant variety and is an F_3-derived selection from the cross Williams x Kingwa (Haun *et al*. 2011). It was released in the United States in 1988. It is a group III maturity cultivar.

5 Genotype of the Modified Plant

As presented above, ICM soybeans are being produced by [ ] CBI-deleted transformation of soybean tissues from non-transgenic cultivar (cv.) Williams 82 using [ ] CBI-deleted]. [ ] is presented in Figure 1. The nucleotide sequence of the DNA [ ] to the soybean genome [ ] is presented below in Figure 2. The order of the genetic components in the [ ], the donor organism of each genetic element, a description of the function of the genes and genetic elements as well as the GenBank number of the DNA sequences of the genes and genetic elements are presented in Table 1.

CBI-deleted CBI-deleted CBI-deleted CBI-deleted CBI-deleted

A [**]** CBI-deleted

Figure 1. [CBI-deleted

] CBI-deleted

[CBI-deleted

CBI-deleted

]

CBI-deleted

[ ] contains CBI-deleted

two gene constructs. The first gene is *lanFP1* encoding the Green Fluorescent Protein (GFP). The second gene transferred to the soybean genome encodes [CBI-deleted

]. Expression of this protein in plant cells [CBI-deleted CBI-deleted

] and serves as a selectable marker for plant CBI-deleted

transformation.

B Sequence of the [ ] insert in ICM soybeans CBI-deleted

Figure 2. Nucleotide sequence of the [ ] insert in ICM soybeans CBI-deleted

1 [ ]AAATT GACGCTTAGA CAACTTAATA ACACATTGCG CBI-deleted

61 GACGTTTTTA ATGTACTGAA TTAACGCCGA ATTAATTCGG GGGATCTGGA TTTTAGTACT

121 GGATTTTGGT TTTAGGAATT AGAAATTTTA TTGATAGAAG TATTTTACAA ATACAAATAC

181 ATACTAAGGG TTTCTTATAT GCTCAACACA TGAGCGAAAC CCTATAGGAA CCCTAATTCC

241 CTTATCTGGG AACTACTCAC ACATTATTAT GGAGAAACTC GAG[CBI-deleted

301 CBI-deleted

361 CBI-deleted

421 CBI-deleted

481 CBI-deleted

541 CBI-deleted

601 CBI-deleted

661 CBI-deleted

721 CBI-deleted

781 CBI-deleted

841 CBI-deleted

901 CBI-deleted

961 CBI-deleted

1021][CBI-deleted CBI-deleted

1081 CBI-deleted

1141 CBI-deleted

1201 CBI-deleted

1261]TGTTTAG TATACTAGAA CBI-deleted

1321 TACCAGCGTG TCTCGAGAGA GATAGATTTG TAGAGAGAGA CTGGTGATTT CAGCGTGTCC

1381 TCTCCAAATG AAATGAACTT CCTTATATAG AGGAAGGTCT TGCGAAGGAT AGTGGGATTG

1441 TGCGTCATCC CTTACGTCAG TGGAGATATC ACATCAATCC ACTTGCTTTG AAGACGTGGT

1501 TGGAACGTCT TCTTTTTCCA CGATGCTCCT CGTGGGTGGG GGTCCATCTT TGGGACCACT

1561 GTCGGCAGAG GCATCTTGAA CGATAGCCTT TCCTTTATCG CAATGATGGC ATTTGTAGGT

1621 GCCACCTTCC TTTTCTACTG TCCTTTTGAT GAAGTGACAG ATAGCTGGGC AATGGAATCC

1681 GAGGAGGTTT CCCGATATTA CCCTTTGTTG AAAAGTCTCA ATAGCCCTTT GGTCTTCTGA

1741 GACTGTATCT TTGATATTCT TGGAGTAGAC GAGAGTGTCG TGCTCCACCA TGTTATCACA

1801 TCAATCCACT TGCTTTGAAG ACGTGGTTGG AACGTCTTCT TTTTCCACGA TGCTCCTCGT

1861 GGGTGGGGGT CCATCTTTGG GACCACTGTC GGCAGAGGCA TCTTGAACGA TAGCCTTTCC

1921 TTTATCGCAA TGATGGCATT TGTAGGTGCC ACCTTCCTTT TCTACTGTCC TTTTGATGAA

1981 GTGACAGATA GCTGGGCAAT GGAATCCGAG GAGGTTTCCC GATATTACCC TTTGTTGAAA

2041 AGTCTCAATA GCCCTTTGGT CTTCTGAGAC TGTATCTTTG ATATTCTTGG AGTAGACGAG

2101 AGTGTCGTGC TCCACCATGT TGGCAAGCTG CTCTAGCCAA TACGCAAACC GCCTCTCCCC

2161 GCGCGTTGGC CGATTCATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG GAAAGCGGGC

2221 AGTGAGCGCA ACGCAATTAA TGTGAGTTAG CTCACTCATT AGGCACCCCA GGCTTTACAC

2281 TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTGAGCG GATAACAATT TCACACAGGA

2341 AACAGCTATG ACCATGATTA CGAATTCCAA TTGTGAGACT TTTCAACAAA GGGTAATATC

2401 CGGAAACCTC CTCGGATTCC ATTGCCCAGC TATCTGTCAC TTTATTGTGA AGATAGTGGA

2461 AAAGGAAGGT GGCTCCTACA AATGCCATCA TTGCGATAAA GGAAAGGCCA TCGTTGAAGA

2521 TGCCTCTGCC GACAGTGGTC CCAAAGATGG ACCCCCACCC ACGAGGAGCA TCGTGGAAAA

2581 AGAAGACGTT CCAACCACGT CTTCAAAGCA AGTGGATTGA TGTGATATCT CCACTGACGT

2641 AAGGGATGAC GCACAATCCC ACTAGTCTTC GCAAGACCCT TCCTCTATAT AAGGAAGTTC

2701 ATTTCATTTG GAGAGAACAC CTAGGCGGCG ATCCACAGGG AAGGAGCAGC ATCTCCACAA

2761 AGACGCACTA CAGAAGACTA AAGAGAGCTT TTTCATACCA AAGAAGTACA ACAAAAGATT

2821 TGCTCCTCAT TTTCTGAATC CTGGGACTCT CTAGCCTGTA GAAGAAGAAA GGCAGGAATT

2881 TCAGCTCAAG AGAACAGATC ACAATATTTA CCCACGGCAC TGTCTCGCAA TCCATGGCAT

2941 TGCCCGCCAC CCATGACATT CATCTGCACG GTTCCATAAA TGGCCACGAG TTCGATATGG

3001 TCGGAGGGGG GAAGGGAGAT CCTAACGCAG GCTCACTGGT AACAACAGCA AAGTCAACTA

3061 AAGGCGCTCT GAAGTTCTCA CCTTACTTGA TGATACCCCA CCTTGGATAC GGGTATTATC

3121 AATATCTTCC ATATCCCGAC GGACCCAGCC CTTTCCAAAC CTCTATGCTT GAAGGCAGTG

3181 GGTATGCTGT GTATCGCGTC TTTGACTTTG AGGACGGGGG AAAGCTCACA ACAGAATTTA

3241 AGTATTCATA CGAAGGCTCA CACATAAAGG CTGACATGAA GTTGATGGGG AGTGGATTCC

3301 CAGACGATGG TCCAGTGATG ACTAGCCAGA TCGTGGACCA GGACGGCTGC GTGAGCAAAA

3361 AGACCTATCT CAATAACAAT ACAATAGTTG ACAGTTTTGA CTGGTCATAT AACTTGCAAA

3421 ACGGAAAAAG ATACCGTGCT CGTGTCAGTT CACATTACAT CTTTGACAAG CCATTTAGTG

3481 CTGATCTCAT GAAAAAACAG CCCGTATTTG TCTACCGCAA ATGCCATGTA AAGGCTAGTA

3541 AGACAGAGGT TACCTTGGAC GAACGTGAGA AGGCATTCTA CGAATTGGCA TGAGGCGCGC

3601 CATATGAAGA TGAAGATGAA ATATTTGGTG TGTCAAATAA AAAGCTTGTG TGCTTAAGTT

```
3661 TGTGTTTTTT TCTTGGCTTG TTGTGTTATG AATTTGTGGC TTTTTCTAAT ATTAAATGAA
3721 TGTAAGATCT CATTATAATG AATAAACAAA TGTTTCTATA ATCCATTGTG AATGTTTTGT
3781 TGGATCTCTT CTGCAGCATA TAACTACTGT ATGTGCTATG GTATGGACTA TGGAATATGA
3841 TTAAAGATAA GCCTGCAGGC ATGCAAGCTT GGCGCGCGCG GTGTCATCTA TGTTACTAGA
3901 TCGGGAATTA AACTATCAGT GTT[                          ]
```

CBI-deleted

C Annotation of the [ ] inserted genetic material CBI-deleted

Table 1. Genetic material inserted into the genome of ICM soybeans

Genetic Element	Position in the genetic insertion	Donor Organism	Function	GenBank No.
[ ]	[ ]	[ ]	[ ]	[ ]
Intervening sequence	26-102	Not applicable. Synthetic sequence, polylinker	Sequence used for DNA cloning.	
CaMV polyadenylation signal	103-277	Cauliflower Mosaic Virus	The 3' Untranslated Region (UTR) of the 35S genes of cauliflower mosaic virus (CaMV) (Mogen *et al*., 1990) that directs polyadenylation in plant cells and terminates transcription of the [ ] gene inserted in the soybean genome.	GenBank: KY703615.1 (7460..7634)
Intervening sequence	278-283	Not applicable. Synthetic sequence, polylinker	Sequence used for DNA cloning.	
[ ]	284-1075	*Escherichia coli*	Bacterial coding sequence [ ] and serves as a selectable marker for plant transformation [ ]	[ ]

CBI-deleted

Table 1 continued. Genetic material inserted into the genome of ICM soybeans

Genetic Element	Position in the genetic insertion	Donor Organism	Function	GenBank No.	
[ ]	1076-1303	*Arabidopsis thaliana*	[ ]	[ ]	CBI-deleted CBI-deleted CBI-deleted CBI-deleted CBI-deleted CBI-deleted CBI-deleted CBI-deleted CBI-deleted
Intervening sequence	1304-1369	Not applicable. Synthetic sequence, polylinker	Sequence used for DNA cloning.		
Enhanced CaMV 35S promoter	1370-2047	Cauliflower Mosaic Virus (CaMV)	The promoter for the 35S genes from the cauliflower mosaic virus (Odell *et al*., 1985) with the duplicated enhancer region (Kay *et al*., 1987) that drives transcription of the [ ] gene.	GenBank: KY703615.1 (7460..7634)	CBI-deleted
Intervening sequence	2048-2373	Not applicable. Synthetic sequence, polylinker	Sequence used for DNA cloning.		
CaMV 35S promoter	2374-2719	Cauliflower Mosaic Virus (CaMV)	The promoter for the 35S genes from the cauliflower mosaic virus (Odell *et al*., 1985) that drives transcription of the *lanFP1* gene.	GenBank: NC_001497.2 (7093..7438)	
Intervening sequence	2720-2725	Not applicable. Synthetic sequence, polylinker	Sequence used for DNA cloning.		

Table 1 continued. Genetic material inserted into the genome of ICM soybeans

Genetic Element	Position in the genetic insertion	Donor Organism	Function	GenBank No.	
OsMac3 5'UTR	2726-2933	*Oryza sativa*	A 5' UTR leader DNA sequence derived from *Oryza sativa* that enhances protein expression in plants. (Aoki *et al*., 2014)	GenBank: NC_029260.1 (25631045.. 25631240)	
lanFP1: Coding sequence for the Green Fluorescent Protein (GFP) from Lancelet	2934-3593	*Branchiostoma floridae*	Synthesis of the GFP is driven by a constitutive promoter and when the GFP receives excitation light it emits green fluorescent light at 509 nm (Baumann *et al*., 2008).	GenBank: XP_035658893.1	
Intervening sequence	3594-3597	Not applicable. Synthetic sequence, polylinker	Sequence used for DNA cloning.		
AtHSP terminator	3598-3874	*Arabidopsis thaliana*	The 3'UTR DNA sequence of the terminator for heat shock protein HSP18.2, that terminates transcription of the *lanFP1* gene (Nagaya *et al*., 2010).	GenBank: NC_003076.8 (24063118.. 24063367)	**CBI-deleted**
Intervening sequence	3875-3923	Not applicable. Synthetic sequence, polylinker	Sequence used for DNA cloning.		
[ ]	[ ]	[ ]	[ ]	[ ]	**CBI-deleted** **CBI-deleted** **CBI-deleted** **CBI-deleted** **CBI-deleted** **CBI-deleted** **CBI-deleted** **CBI-deleted**

6 Description of New Trait

A Intended trait

GFP:
The ICM soybean contains a fluorescence protein that produces a fluorescence signal. The fluorescence trait is only visible when the protein absorbs the correct excitation light and optical filters are used to visualize the fluorescent light.

[ ] **CBI-deleted**
The soybean contains a protein [**CBI-deleted**
] serves as a selectable marker for plant transformation. **CBI-deleted**

B Intended phenotype

GFP:
The phenotype of ICM soybean is green fluorescence with an emission peak at approximately 509 nm.

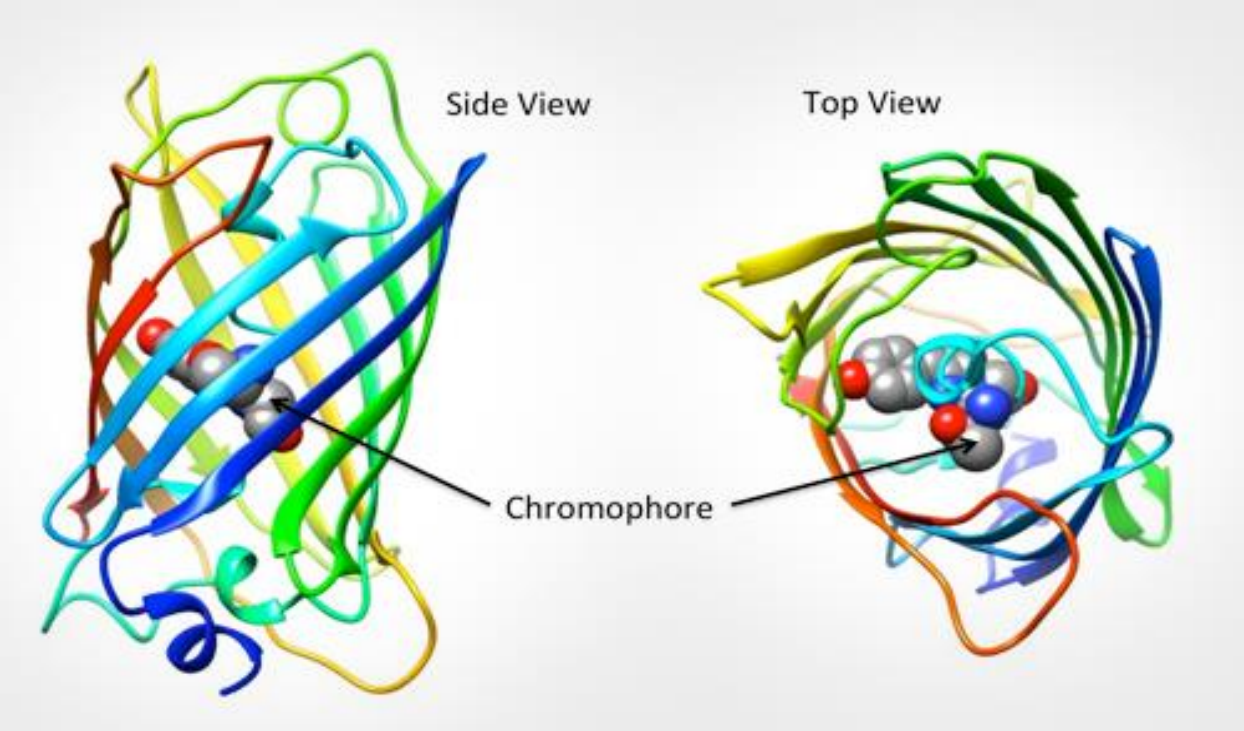
[ ] **CBI-deleted**
ICM soybeans are resistant to the antibiotics [ ] **CBI-deleted**
protein serves as a selectable marker for plant transformation. **CBI-deleted**

C Description of the Mechanism of Action (MOA)

GFP:
The following describes the mechanism of action of GFP.

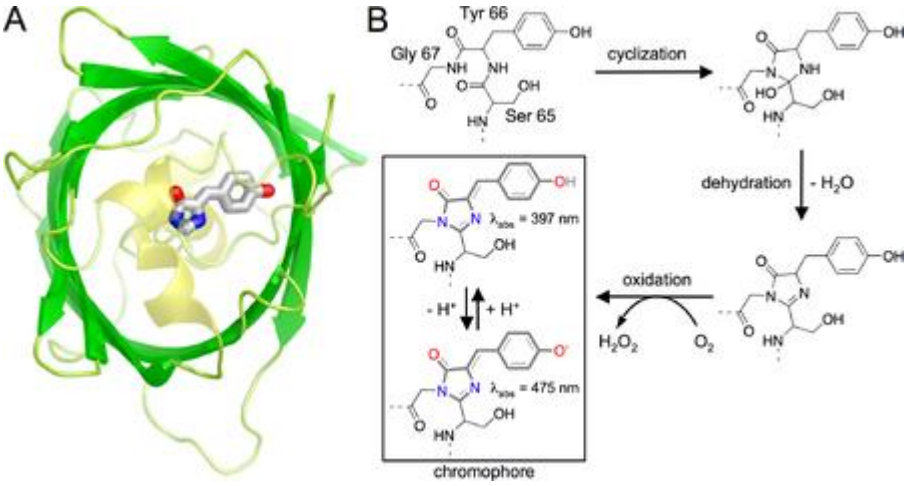
Two independent reports of the x-ray crystal structure of GFP (Ormo *et al*., 1996; Yang *et al*., 1996) revealed that the protein has a unique overall fold comprised of an 11-stranded β-sheet wrapped into a cylindrical β-barrel protein that is 42 amino acids in height and 24 amino acids in diameter (Figure 3). 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 3. 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 4B (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 4. 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).

The safety of GFP has been demonstrated in peer-reviewed literature as well as studies conducted by InnerPlant. Pure GFP and diets containing transgenic canola expressing GFP were fed to young male rats for 26 days to evaluate the potential toxicity and allergenicity 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. Further, the GFP amino acid sequence was analyzed for potential homologies to known protein toxins following the method described in Sharma *et al*. (2022). When the database, ToxinPred2, was searched using default parameters, no significant amino acid sequence homologies between GFP to known protein toxins were found.

The second gene transferred to the soybean genome [CBI-deleted
] Expression of this protein in plant cells [CBI-deleted CBI-deleted
] serves as a selectable marker for CBI-deleted
plant transformation. The following describes the mechanism of action [ ] CBI-deleted

[CBI-deleted
CBI-deleted
CBI-deleted
CBI-deleted
] CBI-deleted

Figure 5. [CBI-deleted
CBI-deleted
] CBI-deleted
[CBI-deleted

CBI-deleted

]

CBI-deleted

[CBI-deleted
CBI-deleted
CBI-deleted
CBI-deleted
] CBI-deleted

[ ] expressed in ICM soybeans is not CBI-deleted
expected to exhibit any enzymatic activity. Therefore, [ ] is not expected to have any effect CBI-deleted
on other soybean metabolic pathways. In addition, the [ ] amino acid sequence was analyzed CBI-deleted
for potential homologies to known protein toxins following the method described in Sharma *et al*. (2022). When the database, ToxinPred2, was searched using default parameters, no significant amino acid sequence homologies between [ ] and known protein toxins were found, CBI-deleted
supporting the safety of the [ ] protein. CBI-deleted

7 Proposed plant-trait-MOA language for website

Plant: Glycine max (soybean)
Trait: Fluorescent marker gene
Phenotype: green fluorescence
MOA: Expression of a fluorescent biomarker

Literature Cited

Aoki, H., Teramura, H., Schepetilnikov, M., and Ryabova, L.A. (2014) Enhanced translation of the downstream ORF attributed to a long 5^|^#8242; untranslated region in the OsMac1 gene family members, OsMac2 and OsMac3. Plant Biotechnology 31(3):221-228.

[**CBI-deleted**

CBI-deleted

] **CBI-deleted**

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.

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.

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

CFIA (1996) The biology of *Glycine max* (L.) Merr. (soybean). Biology Document B101996-10. Canadian Food Inspection Agency, Ottawa, Ontario, Canada. http://www.inspection.gc.ca/english/plaveg/bio/dir/t11096e.shtml

Erickson, E.H. (1975) Variability of floral characteristics influence honey bee visitation to soybean blossom. Crop Sci. 15:767-771.

Erickson, E.H. (1984) Soybean pollination and honey production: A research progress report. Am. Bee J. 145:775-779.

Haun, W.L., Hyten, D.L., Xu, W.W., Gerhardt, D.J., Albert, T.J., Richmond, T., Jeddeloh, J.A., Jia, G., Springer, N.M., Vance, C.P., and Stupar, R.M. (2011) The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. Plant Physiol. 155(2) 645-655.

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.

Hermann, F.J. (1962) A revision of the genus *Glycine* and its immediate allies, USDA Tech. Bull. 1268:1-79.

[
]
CBI-deleted
CBI-deleted

Hymowitz, T. (2004) Speciation and cytogenetics. *In:* Soybeans: Improvement, Production, and Uses. 3rd edition, Boerma, H.R. and Specht, J.E. (Eds.), Agron. Monogr. 16. ASA, CSA, and SSSA Publishers, Madison, WI, USA., pp. 97-136.

Kay, R., Chan, A., Daly, M., and McPherson, J. (1987) Duplication of CaMV35S promoter sequences creates a strong enhancer for plant genes. Science 236(4806):1299-1302.

[
]
CBI-deleted
CBI-deleted
CBI-deleted

Mogen, B.D., MacDonald, M.H., Graybosh, R., and Hunt, A.G. (1990) Upstream sequences other than AAUAAA are required for efficient messenger RNA 3'-end formation in plants. The Plant Cell 2:1261-1272.

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

Nagaya, S., Kawamura, K., Shinmyo, A., and Kato, K. (2010). The HSP terminator of *Arabidopsis thaliana* increases gene expression in plant cells. Plant and Cell Physiol. 51(2):328-332.

Odell, J.T., Nagy, F., and Chua, N.H. (1985). Identification of DNA sequences required for activity of the Cauliflower Mosaic Virus 35S promoter. Nature 313:810-812.

OECD (2000) Organization for Economic Cooperation and Development, Consensus document on the biology of *Glycine max* (L.) Merr. (Soybean). Series on Harmonization of Regulatory Oversight in Biotechnology No. 15. ENV/JM/MONO(2000)9. http://www.oecd.org.

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.

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

Raper, C.D., and Kramer, P.J. (1987) Stress physiology. In: Soybeans: Improvement, Production, and Uses. 2nd ed., Wilcox, J.R. (Ed.), ASA, CSSA, SSSA, Madison, WI., pp 589-641.

Ray, J.D., Kilen, T.C., Abel, A.C., and Paris, R.L. (2003) Soybean natural cross-pollination rates under field conditions. Environ. Biosafety Res. 2:133-138.

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

Shoemaker, R.C., Hatfield, P.M., Palmer, R.G., and Atherly, A.G. (1986) Chloroplast DNA variation in the genus *Glycine* subenus *Soja*. J. Hered. 77:26-30.

Singh, R.J., and Hymowitz, T. (1989) The genomic relationships among *Glycine soja* Sieb. and Zucc., *G. max* (L.) Merr. and '*G. gracilis*' Skvortz. Plant Breed. 103:171-173.

[

CBI-deleted

CBI-deleted

]

CBI-deleted

TeKrony, D.M., Egli, D.B., and White, G.M. (1987) Seed production and technology. In Soybeans: Improvement, Production, and Uses. 2nd ed. Wilcox, J.R. (Ed.), American Society of Agronomy. Madison, WI., pp. 295-354.

USDA. (2006) USDA federal noxious weed list. http://www.aphis.usda.gov/plant_health/plant_pest_info/weeds/downloads/weedlist2006.pdf.

USDA-APHIS. (2008) http://www.aphis.usda.gov/brs/soybean.html.

Wang, C.L. (1976) Review on the classification of soybeans. (In Chinese) Acta Phytotaxon. Sin. 14:22-30.

[

CBI-deleted

CBI-deleted

]

CBI-deleted

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.