ORIGINAL SUBMISSION

Morgan, Lewis & Bockius LLP
1111 Pennsylvania Avenue, NW
Washington, DC 20004
Tel. 202.739.3000
Fax: 202.739.3001
www.morganlewis.com

Morgan Lewis
COUNSELORS AT LAW

Gary L. Yingling
Partner
202.739.5610
gyingling@morganlewis.com

September 4, 2014

VIA FEDERAL EXPRESS

Dr. Antonia Mattia
Director
Division of Biotechnology and GRAS Notice Review
Office of Food Additive Safety (HFS-200)
Center for Food Safety and Applied Nutrition
Food and Drug Administration
5100 Paint Branch Parkway
College Park, MD 20740-3835

Re: GRAS Notification for Soybean Leghemoglobin Protein Derived from *Pichia pastoris*

Dear Dr. Mattia:

On behalf of Impossible Foods Inc. ("Impossible Foods"), we are submitting under cover of this letter three paper copies and one eCopy of Impossible Foods generally recognized as safe ("GRAS") notification for its soybean leghemoglobin protein derived from *Pichia pastoris*. The electronic copy is provided on a virus-free CD, and is an exact copy of the paper submission. Impossible Foods has determined through scientific procedures that its soybean leghemoglobin protein produced by a submerged batch fermentation of *Pichia pastoris* is GRAS for use as a component of meat and poultry analogue products.

The soybean leghemoglobin protein is to be used as a plant-based protein component in non-animal derived food products with the texture, nutrition, flavor and appearance of traditional animal derived foods. This modified soy protein will impart a flavor impact to meat analogue products, and enhance the dietary profile of those products.

Pursuant to the regulatory and scientific procedures established by proposed regulation 21 C.F.R. § 170.36, this use of soybean leghemoglobin protein derived from *Pichia pastoris* is exempt

Morgan, Lewis & Bockius LLP
1111 Pennsylvania Avenue, NW
Washington, DC 20004
Tel. 202.739.3000
Fax: 202.739.3001
www.morganlewis.com

Morgan Lewis
COUNSELORS AT LAW

GRN 000540

Gary L. Yingling
Partner
202.739.5610
gyingling@morganlewis.com

September 4, 2014

VIA FEDERAL EXPRESS

Dr. Antonia Mattia
Director
Division of Biotechnology and GRAS Notice Review
Office of Food Additive Safety (HFS-200)
Center for Food Safety and Applied Nutrition
Food and Drug Administration
5100 Paint Branch Parkway
College Park, MD 20740-3835

Re: GRAS Notification for Soybean Leghemoglobin Protein Derived from *Pichia pastoris*

Dear Dr. Mattia:

On behalf of Impossible Foods Inc. ("Impossible Foods"), we are submitting under cover of this letter three paper copies and one eCopy of Impossible Foods generally recognized as safe ("GRAS") notification for its soybean leghemoglobin protein derived from *Pichia pastoris*. The electronic copy is provided on a virus-free CD, and is an exact copy of the paper submission. Impossible Foods has determined through scientific procedures that its soybean leghemoglobin protein produced by a submerged batch fermentation of *Pichia pastoris* is GRAS for use as a component of meat and poultry analogue products.

The soybean leghemoglobin protein is to be used as a plant-based protein component in non-animal derived food products with the texture, nutrition, flavor and appearance of traditional animal derived foods. This modified soy protein will impart a flavor impact to meat analogue products, and enhance the dietary profile of those products.

Pursuant to the regulatory and scientific procedures established by proposed regulation 21 C.F.R. § 170.36, this use of soybean leghemoglobin protein derived from *Pichia pastoris* is exempt

from premarket approval requirements of the Federal Food, Drug and Cosmetic Act, because the notifier has determined that such use is GRAS.

If you have any questions regarding this notification, or require any additional information to aid in the review of Impossible Foods's conclusion, please do not hesitate to contact me via email at gyingling@morganlewis.com or by telephone, (202)739-5610.

Sincerely,

(b) (6)

Gary L. Yingling

cc: Impossible Foods Inc.

GRAS NOTIFICATION FOR SOYBEAN LEGHEMOGLOBIN PROTEIN DERIVED FROM *PICHIA PASTORIS*

Submitted by:
Impossible Foods Inc.
525 Chesapeake Drive
Redwood City, CA 94063

TABLE OF CONTENTS

1. GENERAL INTRODUCTION AND CLAIM OF EXEMPTION FROM PREMARKET APPROVAL REQUIREMENTS 1
1.1 Name and Address of Notifier 1
1. 2 Common or Usual Name of Soy Leghemoglobin Protein 1
1. 3 Applicable Conditions of Use 2
1. 4 Basis for GRAS Determination 2
1. 5 Availability of Information for FDA Review 3
2. IDENTITY OF MODIFIED SOY PROTEIN 4
2.1. Chemical Name 4
2.2. Composition 4
2.3. Specification of RUBIA 4
2.4 Method of Manufacture 6
2.4. 1 Preparation of the Production Strain for Fermentation 6
2.4. 2 Method of Manufacture 8
3. SELF-LIMITING LEVELS OF USE 10
4. BASIS FOR GRAS DETERMINATION 10
4.1. History of Safe Use 10
4.2. Probable Consumption/Estimated Dietary Intake 12
4.3 Estimated Daily Intake of Modified Soy Protein 13
a. Estimation of "90th Percentile" Intake 14
4.4 Assessment of Allergenicity and Toxicology of Soy Leghemoglobin 14
4.5 Expert Panel Report 15
4.5 GRAS Conclusion 31
LIST OF REFERENCES 33
LIST OF ANNEXES 36

1. GENERAL INTRODUCTION AND CLAIM OF EXEMPTION FROM PREMARKET APPROVAL REQUIREMENTS

1.1 Name and Address of Notifier

NOTIFIER

Impossible Foods Inc.
525 Chesapeake Drive
Redwood City, CA 94063
Phone: (650) 461-4385

MANUFACTURER

Impossible Foods Inc.
525 Chesapeake Drive
Redwood City, CA 94063
Phone: (650) 461-4385

PERSON RESPONSIBLE FOR THE DOSSIER

Mr. Nick Halla
Impossible Foods Inc.
525 Chesapeake Drive
Redwood City, CA 94063
Phone: (650) 461-4381

1. 2 Common or Usual Name of Soy Leghemoglobin Protein

As discussed in greater detail in other sections of this report, Impossible Foods Inc. ("Impossible Foods") has determined that the use of soy leghemoglobin protein as a component of meat and poultry analogue products is generally recognized as safe. Impossible Foods will market soy leghemoglobin protein in the United States under the trade name RUBIA.

According to the section 403(I)(2) of the Federal Food, Drug, and Cosmetic Act and 21 C.F.R. § 101.4, a food product must list the common or usual name of each ingredient in the food. Impossible Foods recognizes, in this GRAS notification, that an appropriate common or usual name of soy leghemoglobin protein is "modified soy protein."

1. 3 Applicable Conditions of Use

Modified soy protein is to be used as a plant-based protein component in non-animal derived food products with the texture, nutrition, flavor and appearance of traditional animal derived foods. Modified soy protein will impart a flavor impact to meat analogue products. Thus the use of modified soy protein in meat analogue products will enhance the dietary profile of those products (Carpenter & Mahoney, 1992) (Proulx & Reddy, 2006).

Modified soy protein, along with several other Food and Drug Administration ("FDA" or "Agency") approved plant proteins, will be components of the meat and poultry analogue products. Other proteins may include, but are not limited to, commercially available proteins from soy, pea, mung bean, lentil, corn, and wheat. Modified soy protein will function to contribute to the flavor, appearance and nutritional quality of meat analogue products. A typical meat analogue product will contain:

Component	Meat Analogue
Protein	10%-25%
Plant Oils	0%-25%
Miscellaneous+	2%
Water	60%-75%

+Miscellaneous ingredients include salt, flavors, spices and seasonings, emulsifiers, etc.

RUBIA will be added to the meat or poultry analogue product to deliver not more than 1% modified soy protein. RUBIA is formulated to contain 73% modified soy protein on a solids basis. Thus, RUBIA will constitute not more than 1.4% of the total composition of meat and poultry analogue products in which it is used. The use of modified soy protein in meat and poultry analogue products above the specified use-levels is largely self-limiting based on unacceptable organoleptic properties.

1. 4 Basis for GRAS Determination

Pursuant to the regulatory and scientific procedures set forth in the Proposed Rule "Substances Generally Recognized as Safe," 62 Fed. Reg. 18937 (April 17, 1997) (proposed 21 C.F.R. § 170.36) ("GRAS Proposed Rule"), Impossible Foods has determined, through scientific procedures, that its soy leghemoglobin protein derived from soybean and expressed in *Pichia pastoris* is a GRAS substance for the intended food applications and is therefore exempt from the requirement for premarket approval.

An independent panel of recognized experts qualified by their individual scientific training and experience, using appropriate scientific procedures, evaluated the safety of modified soy protein under the conditions of its intended use in food as a component of meat and poultry replacement products (meat analogues).

The panel assessed the origin, production, and intended use of soybean leghemoglobin for use in meat and poultry replacement (analogue) products. The panel also considered the potential allergenicity and toxicity of soybean leghemoglobin and the chemical and functional equivalency of soybean leghemoglobin with plant and animal hemoglobin proteins commonly consumed in the diet.

The panel unanimously concluded that soybean leghemoglobin protein is Generally Recognized As Safe (GRAS) based on scientific procedures (structural and functional equivalency to other safe plant and animal hemoglobins).

1. 5 Availability of Information for FDA Review

The data and information that are the basis for GRAS determination are available for the FDA's review and copies will be sent to FDA upon request. Requests for copies and arrangements for review of materials cited herein may be directed to:

Mr. Gary Yingling
Morgan, Lewis & Bockius LLP
1111 Pennsylvania Avenue, NW
Washington, D.C. 20004-2541
Phone: (202) 739-5610

2. IDENTITY OF MODIFIED SOY PROTEIN

The data and information in this notification establishes that hemoglobins, including soy hemoglobin proteins, are widely consumed in the diet through the consumption of traditional foods and the proposed additional usage of soy leghemoglobin protein in meat and poultry replacement products will not significantly increase that exposure.

2.1. Chemical Name

The chemical name of the characterizing component of modified soy protein is soy leghemoglobin (UniProtKB/Swiss-Prot #: P02236). The source of the protein is the soybean plant *Glycine max* gene *LGB2*.

2.2. Composition

Hemoglobin proteins are found in most organisms, including bacteria, protozoa, fungi, plants and animals (Hardison, 1998). Hemeproteins are classified as globin/non-globin and symbiotic/non-symbiotic. Hemoglobin, myoglobin, and leghemoglobin are examples of globin proteins. Cytochrome oxidases, hemocyanins, and methemalbumin are examples of non-globin hemeproteins (Everse, 2004; Jokipii-Lukkari, 2009). Plant hemoglobins are classified according to function as symbiotic or non-symbiotic (Gupta, 2011). Symbiotic hemoglobins are found predominantly in legumenous plant species. The most studied symbiotic hemoglobins are the leghemoglobins of nitrogen fixing legumes where they facilitate oxygen diffusion within root tissues. Nonsymbiotic hemoglobins have been identified in a wide range of legume and nonlegume plants. The highest expression levels for nonsymbiotic plant hemoglobin are observed in metabolically active or stressed tissue (Anderson C. R., 1996).

Impossible Foods has analyzed hemoglobin sequences from various sources (including corn, rice, soy, barley, lupine, horse, tuna, and pig myoglobin). As detailed in Annex 1, these hemeproteins (animal myoglobins, plant hemoglobins and plant leghemoglobins) are structurally very similar and contain an identical heme B cofactor.

Impossible Foods enlisted Dr. Richard E. Goodman at the Food Allergy Resource and Research Program (FARRP) of the University of Nebraska to assess the potential allergenicity and toxicity of soy leghemoglobin, as discussed in Section 4.3. Dr. Goodman's assessment included a full literature search to identify any published literature regarding possible allergenicity or toxicity associated with hemoglobin proteins. The conclusion of this assessment was that no published literature could be found that suggested allergic, toxic or adverse health effects related to consumption of any hemoglobin proteins. Analysis of the amino acid sequence of soy leghemoglobin revealed that soy leghemoglobin does not demonstrate sufficient similarity with any known allergen or toxin to raise concern (Annex 3).

2.3. Specification of RUBIA

Following purification, RUBIA is standardized to contain at least 80 grams per liter (g/l) soy leghemoglobin protein. RUBIA is stabilized with 20 mM potassium phosphate and 100 mM

sodium chloride. All stabilizing agents are food grade. The product specification of RUBIA is presented in Table 1. RUBIA may be stored at -20 °C as a frozen liquid or dried powder for 12 months with no observable change in leghemoglobin stability or performance in meat and poultry analogue products.

Table 1. Specifications of RUBIA

	Concentration (Frozen Liquid) (% w/w)	**Concentration (Dry Powder) (% w/w)**
Protein	10	91
Leghemoglobin	8	73
Ash	<1	<9
Fat	<0.1	<1
Carbohydrate	<0.1	<1
Solids	11	100
Moisture	89	0

	Concentration (Frozen LIquid) (ppm w/w)	**Concentration (Dry Powder) (ppm w/w)**
Lead	<0.01	<0.1
Arsenic	<0.01	<0.1
Mercury	<0.005	<0.05
Cadmium	<0.1	<1

	Concentration (Frozen Liquid)	**Concentration (Dry Powder)**
Aerobic plate count[1] (CFU/g)	<10^4	<10^4
E. coli O157H7[2]	Absent by test	Absent by test
Salmonella spp[3]	Absent by test	Absent by test
Listeria monocytogenes[4]	Absent by test	Absent by test

[1] AOAC OMA 990.12
[2] AOAC RI 020801
[3] AOAC OMA 2011.03
[4] AOAC OMA 2010.02

Three lots of RUBIA were analyzed for chemical and microbial composition to demonstrate that all three lots meet the specifications listed above (Annex 5).

2.4 Method of Manufacture

RUBIA is prepared in four stages: construction of the production strain of *Pichia pastoris*, expression of soy leghemoglobin protein in submerged fermentation, purification and stabilization of the expressed soy leghemoglobin protein. All materials used in the production of RUBIA are standard food grade or pharmaceutical grade ingredients of a purity and quality suitable for their intended use (Aunstrup, Andersen, Falch, & Nielsen, 1979) (Taylor & Baumert, 2013) (Enzyme technical association, 2005) and processing conditions are appropriate for food production under cGMP.

2.4. 1 Preparation of the Production Strain for Fermentation

a. Recipient microorganism

This section was written and prepared by BioGrammatics Inc. (Carlsbad, CA) to describe the lineage of their commercially available Bg10 *Pichia pastoris* strain.

The recipient microorganism is *Pichia pastoris* Bg10. The general taxonomy of *P. pastoris* is as follows:
Name: *Pichia pastoris*
Kingdom: Fungi
Phylum: Ascomycota
Class: Hemiascomycetes
Order: Saccharomycetales
Family: Endomycetaceae
Genus: Pichia
Species: pastoris

The production *Pichia pastoris* strain Bg10 was derived from the well characterized strain Y-11430, which is deposited in the collection at the Northern Regional Research Laboratories (NRRL). The lineage of *P. pastoris* strain NRRL Y-11430 is detailed below, and was previously included in GRN 204, reviewed by the Agency in 2006.

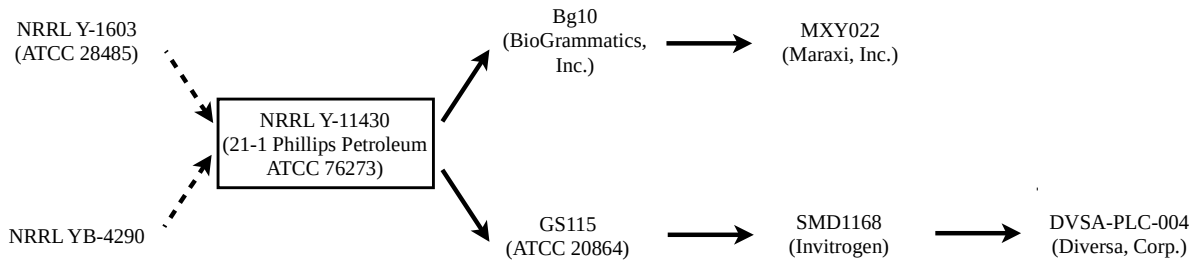
According to the definitive source of yeast taxonomy (Rij, 1984) as well as a thorough literature search, there are no indications that *P. pastoris* has been associated with animal or human illness. The following lineage for the *P. pastoris* Bg10 strain is based on genomic sequencing, literature sources, and from discussions with experts in this area. The first *P. pastoris* strains were isolated from an oak tree and a chestnut tree and were deposited in the collection at the Northern Regional Research Laboratories (NRRL)[1] (see Figure 1, and www.biogrammatics.com). Yeast strains screened by Phillips Petroleum for growth on methanol included two *P. pastoris* strains, designated NRRL Y-1603 (ATCC accession 28485) (ATCC, 2006b) and NRRL YB-4290 (NCAUR, 2006). Phillips Petroleum identified a *P. pastoris* strain with improved growth characteristics. The strain was designated 21-1 and deposited at NRRL, as

[1] The NRRL collection is now known as the Agriculture Research Service Culture Collection and is at the Microbial Genomics and Bioprocessing Research Unit (MGB) of the National Center for Agricultural Utilization Research (NCAUR) in Peoria, IL.

NRRL Y-11430 (Wegner, E.H., 1986). This strain is now available from ATCC as 76273 (ATCC, 2005). No records are available confirming that NRRL Y-1603 or NRRL YB-4290 is the progenitor of NRRL Y-11430, but it seems likely that one of them is the progenitor strain (Madden, K.R., 2014). NRRL Y-11430 was the progenitor strain for GS115, a histidine auxotrophic mutant *(his4-)* (ATCC, 2006a; (Cregg, 1985)), a common *Pichia* strain provided in commercial kits by Invitrogen Corporation, and widely used as the parental strain of many biotechnology products, including FDA approved proteins such as Kalibitor® (ecallantide, for the treatment of acute attacks of hereditary angioedema, 2009). Additionally, the GS115 derived strain SMD1168 is used for the GRAS approved production of BD16449 Phospholipase C (GRN 204). Like GS115, the BioGrammatics, Inc. strain, Bg10 is also a derivative of NRRL Y-11430, and genomic sequencing data performed by BioGrammatics Inc. confirm the similarity of NRRL Y-11430, Bg10 and GS115 (Figure 1). Additional taxonomic history of these strains is available in a 2009 manuscript by C. Kurtzman (Kurtzman, 2009) and on the Biogrammatics webpage (biogrammatics.com).

BioGrammatics, Inc. further developed the NRRL-Y-11430 strain to remove the native *P. pastoris* plasmids using PCR primers unique to the plasmids to screen multiple single-colony isolates for the presence of the plasmids. One isolate without plasmids was selected to become the wild-type (wt) BioGrammatics strain, Bg10. Genomic sequence from Bg10 indicates the plasmids are no longer present, and, benchmarks the similarity of Bg10 with NRRL-Y11430, as well as with GS115. Like NRRL Y-11430 and GS115, Bg10 does not contain antibiotic resistance genes.

Figure 1. Strain lineage of Impossible Foods (formerly Maraxi) production strain MXY022.



b. Expression vector

The protein coding sequence from *Glycine max* leghemoglobin *LGB2* was synthesized, codon-optimized for expression in *Pichia pastoris,* and sub-cloned into the commercially available expression vector *pJAN* (BioGrammatics, Inc.) to generate *pJAN-legH*. The plasmid map of *pJAN* is publically available at biogrammatics.com. All components of *pJAN-legH* were confirmed by sequencing. The alcohol oxidase promoter (*pAOX1*) and terminator regulate LegH expression. This promoter has been demonstrated to produce high levels of recombinant proteins after producing biomass on glycerol, and inducing *pAOX1* with methanol. *pJAN-legH* contains two antibiotic resistance genes AMP^R (beta-lactamase) and NAT^R, which were used only for

strain construction purposes[2]. Antibiotics were not used during fermentation and RUBIA does not contain the antibiotic resistance genes as demonstrated by quantitative PCR (qPCR) with primers targeting the AMP^R and NAT^R genes. No DNA was observed at a level of detection of 0.001 pg of DNA (Annex 5). Typical yeast transformations require microgram quantities of recombinant DNA for homologous integration; this is approximately 1,000,000,000 times more than the level of detection of this qPCR method. Furthermore, RUBIA does not transform chemically competent *E. coli* cells to acquire AMP or NAT resistance (Annex 5). RUBIA does not contain the production strain MXY022, as demonstrated by the absence of the formation of nourseothricin (NAT) resistant colonies when 0.5 ml of RUBIA is plated on growth medium containing 50 ug/ml NAT and grown for 72 hrs at 30 degrees Celsius (Annex 5). On the rare occasion that a NAT-resistant colony is observed, the colony is analyzed microscopically for *Pichia pastoris* morphology and colony PCR is performed to ensure the absence of MXY022 genomic DNA.

c. Recombinant production strain

P. pastoris Bg10 was transformed with linearized *pJAN-legH* to generate the production strain MXY22 (Figure 1). Transformed cells were isolated and characterized by both colony PCR and quantitative PCR (qPCR) to confirm that a single copy of the *pJAN-LegH* plasmid was integrated into the genome of Bg10 cells at the *AOX1* promoter.

2.4. 2 Method of Manufacture

i. Raw Materials

Raw materials used in the fermentation and recovery process for leghemoglobin are standard ingredients used in the food/enzyme industry and follow internal specifications (in line with Foods Chemical Codex, Ninth Edition requirements). These specifications include limits on lead and other pertinent heavy metals. The raw materials are of a purity and quality suitable for their intended use (Aunstrup, Andersen, Falch, & Nielsen, 1979); they are food grade and GRAS, or high-quality chemical or pharmaceutical grades (USP, NF, or ACS grades) from approved suppliers.

ii. Fermentation

Soy leghemoglobin protein is expressed during submerged fed-batch fermentation using the *P. pastoris* MXY022 production strain described above (Figure 1). Frozen cell banks for the production organism MXY022 are maintained at -80 °C in 15% v/v glycerol as the source inoculum for leghemoglobin production. The master cell bank is stored at multiple locations. Working cell banks are prepared from the master cell bank and are tested for microbial purity, specific growth rate, and leghemoglobin yield prior to

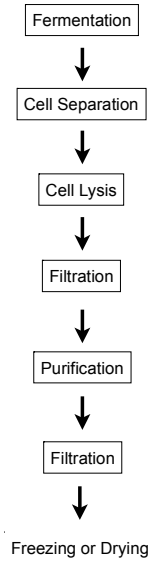
[2] Antibiotic resistance genes were used to ease the burden of strain construction. They will be removed from our final commercial strain.

fermentation. Fermentation broth is periodically analyzed microscopically to ensure culture purity. Process parameters including pH, temperature, agitation, dissolved oxygen, methanol concentration and glycerol concentration are routinely monitored throughout fermentation. Fermentations that incur microbial contamination and/or process deviations that affect safety and/or quality are sterilized by steam in place and discarded.

iii. Recovery Process

The *P. pastoris* cells in the fermentation broth are concentrated via centrifugation and then lysed by high-pressure homogenization. Insoluble material within the lysate is removed by filtration. Soy protein is purified from the clarified lysate by chromatographic separation. Ultrafiltration is used to concentrate soy leghemoglobin protein up to 80 g/l. The resulting concentrated sample is stored either as a frozen liquid or a dehydrated powder. A schematic overview of the manufacturing process is presented in Figure 2.

Figure 2. Schematic overview of the manufacturing process for soy leghemoglobin protein



Impossible Foods' QA/QC Department tests cell paste from each independent fermentation broth to ensure the absence of *Salmonella* AOAC OMA 2011.03, *Listeria monocytogenes* AOAC OMA 2010.02, and *E. coli* O157:H7 AOAC RI 020801. The final RUBIA product from every production run is tested for total aerobic plate count AOAC OMA 990.12 and *Salmonella*, *Listeria monocytogenes*, and *E. coli* O157:H7 as described above. The presence of a pathogen or >10^4 CFU/g would result in the batch being discarded, the execution of additional sanitization standard operating procedures (SSOPs) in compliance with Impossible Food's internal food-safety regulatory standards, and a root cause analysis.

3. SELF-LIMITING LEVELS OF USE

The use of leghemoglobin in meat and poultry analogue products is largely self-limiting because use rates above 1.4% RUBIA results in an unacceptable organoleptic profile.

4. BASIS FOR GRAS DETERMINATION

The data and information in this notification establish that hemoglobins, including soy hemoglobin proteins, are widely consumed in the diet through the consumption of traditional foods and the proposed additional usage of soy leghemoglobin protein in meat and poultry replacement products will not significantly impact current exposure levels.

4.1. History of Safe Use

The safety of soy protein is well established (Riaz, 2006). Soybeans have been part of the human diet for more than 5000 years (Lee, Crawford, Liu, Sasaki, & Chen, 2011). In the 2010 marketing year, 249 million metric tons of soybeans were produced worldwide (Food and Agricultural Organization of the United Nations, 2010). Although the majority of the crop is used for animal feed, approximately 14% is used for human food in the form of traditional soy foods, such as tofu, soymilk, natto, miso, and bean sprouts. Soy protein ingredients are also used to formulate a wide range of food products, including infant formula, dairy and meat alternatives, nutritional supplements and energy bars (Golbitz & Jordan, 2006). The use of soy proteins is widely accepted in the United States. The FDA has affirmed the safety of soy protein isolates for inclusion in many products (GRN 134, GRN 186, and GRN 283) (Food and Drug Administration, 2004) (Food and Drug Administration, 2006) (Food and Drug Administration, 2009) (Food and Drug Administration, 2010), and has approved a health claim for soy protein and the reduced risk of coronary heart disease (21 CFR 101.82). In 2000, the U.S. Department of Agriculture (USDA) issued a ruling allowing soy protein to completely replace animal protein in the National School Lunch Program (Messina, 2006). Thus, the safety of soy in human food has been clearly demonstrated and its use reviewed extensively by United States regulatory agencies.

Hemeproteins are found in most organisms, including bacteria, protozoa, fungi, plants and animals (Everse, 2004) (Hardison, 1998) (Wajcman & Kiger, 2002). Soybeans have been shown to express three hemoglobin proteins: symbiotic, nonsymbiotic and truncated (Lee, Kim, & An, 2004). Symbiotic hemoglobins, found predominately in legume species, function in the nitrogen fixation process in concert with the bacterium *Rhizobium* where they facilitate oxygen diffusion within host tissues. Symbiotic plant hemoglobins, which evolved from non-symbiotic hemoglobins (Gupta, 2011) (Wajcman & Kiger, 2002), are commonly referred to as leghemoglobins. Leghemoglobins' structure and oxygen binding mechanism are similar to those of animal muscle myoglobin proteins (Hargrove, 1997).

Non-symbiotic plant hemoglobins from soybeans, barley, rice, corn, and mung beans are readily consumed in the diet. Anderson, et.al. demonstrated that the nonsymbiotic hemoglobin in soybeans was expressed in various plant tissues including stems, shoots, cotyledon, leaves, and root hair (Anderson C. R., 1996). These soybean tissues are commonly consumed in the diet in

the form of bean sprouts. Sprouted barley, which is widely used in the beverage industry (malted barley) and in the baking industry (malted barley flour), has been shown to express hemoglobin 1 day after imbibition (Duff, Guy, Xianzhou, Durnin, & Hill, 1998). Non-symbiotic hemoglobins are expressed in the rice embryo as well as in the coleoptiles and seminal root of sprouted rice, which is consumed as part of the diet. (Lira-Ruan, Ruiz-Kubli, & Arredondo-Peter, 2011)

Non-symbiotic hemoglobin is expressed in corn seedlings and may provide a good source of bioavailable heme in mature corn seeds (Bodnar, 2011). Impossible Foods has identified the presence of non-symbiotic hemoglobin in mung bean sprout purchased at a local market (Annex 6). The three dimensional structure of soybean leghemoglobin is highly similar to the non-symbiotic hemoglobins of corn, rice, and barley (Annex 1). Although there are no crystal structures for non-symbiotic hemoglobins from soybean or mung beans, based on the highly similar structures of non-symbiotic hemoglobins from corn, rice and barley to each other and soybean leghemoglobin, we expect that they will likewise be similar to leghemoglobin (Annex 1).

As noted previously, symbiotic hemoglobins are predominately found in the nodules of nitrogen fixing plants. There are clear indications that these proteins are consumed. Legumes (alfalfa and peanuts, for example) are common forage crops used to graze sheep, cattle and horses. These animals often consume the entire plant as they graze, including the leaf, stem and root. Additionally, there are references to the use of peanut root, consumed by certain populations in Asia, as a source of protein, minerals and antioxidants (Schanz).

Thus, it is clear that hemoglobin proteins of plant and animal sources are widely consumed in the human diet, and represent a highly bioavailable source of dietary iron for human nutrition. Proulx and Reddy demonstrated that soybean leghemoglobin and bovine hemoglobin showed similar iron bioavailability within a food matrix, both of which were higher than free iron. (Proulx & Reddy, 2006) Furthermore, plant-derived hemoglobins are already prevalent in our food system through malted grain products and sprouted seeds, grains, rice and beans (pulses). (Anderson, Jensen, Leewellyn, Dennis, & and Peacock, 1996) (Duff, Guy, Xianzhou, Durnin, & Hill, 1998)(Lira-Ruan, Ruiz-Kubli, & Arredondo-Peter, 2011)

As discussed in greater detail in other sections of this report, soy leghemoglobin protein is produced in the well-characterized expression system *Pichia pastoris* (Cereghino & Cregg, 2000). This genetically modified production strain complies with the OECD (Organization for Economic Development) criteria for GILSP (Good Industrial Large Scale Practice) microorganisms. (OECD, 1992) It also meets the criteria for a safe production microorganism as described by Pariza and Foster, Pariza and Johnson, and several expert groups (Berkowitz & Maryanski, 1989) (EU Scientific Committee for food, 1992) (FAO/WHO, 1996) (International Food Biotechnology Council, 1990) (Jonas, et al., 11996) (OEDC, 1993) Pariza, M.W. et al., 1983; Pariza, M.W. et al., 2001). The American Association of Feed Control Officials (AAFCO) has approved the *E. coli* enzyme phytase derived from the fermentation of recombinant *Pichia pastoris* for use in animal feed (AAFCO, 2013). Many therapeutic proteins have been recombinantly expressed in *Pichia pastoris* for pharmaceutical use. Angiostatin and Endostatin completed phase I clinical trails and have entered phase II clinical trials (Herbst, et al., 2002). *P. pastoris* is also approved

by FDA as an animal feed protein source allowed in broiler feed up to 10% of the total feed (FDA 21 CFR Part 573, 1993).

Pichia pastoris does not produce active toxins (Pariza & Johnson, 2001). *Pichia pastoris* has been placed in the Biosafety Level 1 (BL-1) class by the ATCC organization, indicating Pichia is a well-characterized agent not known to cause disease in healthy human adults and to be of minimal hazard to laboratory personnel and the environment (Center for Disease Control, 1999). Toxicity studies done in support of the above referenced *P. pastoris*-approved animal feed (including a pathogenicity study in mice, an acute oral toxicity study in rats, a subacute oral toxicity study in rats, and a two generation teratology study in rats) also demonstrated that *P. pastoris* is neither pathogenic nor toxigenic (FDA 21 CFR Part 573, 1993).

The Impossible Foods' *Pichia pastoris* production strain MXY022 is derived from a strain lineage with a long history of safe use, as outlined in GRN 204. All genetic modifications made to generate MXY022 are well characterized and conform to the guidelines for generating safe production strains for the recombinant production of food ingredients (Olempska-Beer, Merker, Ditto, & DiNovi, 2006). RUBIA does not contain the production organism, transformable recombinant DNA, or antibiotic resistance genes (Annex 5).

All materials used in the production of RUBIA are standard food grade or pharmaceutical grade ingredients used in the food industry. The raw materials are of a purity and quality suitable for their intended use (Aunstrup, Andersen, Falch, & Nielsen, 1979). The process follows current Good Manufacturing Practices (cGMP) to isolate soy leghemoglobin protein from a well characterized fermentation medium that complies with the Enzyme Technical Association's guidelines for microbially derived recombinant proteins (Enzyme Technical Association, 2005) (Taylor & Baumert, 2013). The product is standardized to a concentration of at least 80 g/l soy leghemoglobin protein. The soy leghemoglobin protein has highly advantageous properties in meat and poultry replacement products, which will provide consumers a nutritious and flavorful alternative to foods derived from animal products.

4.2. Probable Consumption/Estimated Dietary Intake

The vast majority of hemeproteins consumed in the diet are as myoglobin through consumption of meat and poultry products. Hemeprotein consumption was estimated using data from the "Retail Commodity Intakes: Mean Amounts of Retail Commodities per Individual, 2007-08. (Bowman, Martin, Clemens, Lin, & Moshfegh, 2013) For the US population, *per capita* mean consumption of meat and poultry products is 154 g/day. Following the FDA guidance document "Guidance for Industry: Estimating Dietary Intake of Substances in Food" to estimate daily intake values,[3] the pseudo 90th percentile for meat and poultry consumption would

[3] Available at: http://www.fda.gov/food/guidanceregulation/guidancedocumentsregulatoryinformation/ingredientsadditivesgraspackaging/ucm074725.htm

be 2 times the mean EDI or 308 g/person/day. Assuming an average myoglobin concentration for meat and poultry products of 0.5% (Yip & Dallman, 1996), the average *per capita* myoglobin consumption would be 0.77 g/day and the 90th percentile intake would be 1.54 g/day.

4.3 Estimated Daily Intake of Modified Soy Protein

RUBIA will be marketed for use in meat and poultry analogue products that provide consumers a flavorful and nutritious alternative to meat containing products. Therefore, Impossible Foods has estimated daily intakes of RUBIA by assuming consumers will substitute the meat analogue product for the traditional meat product on a 1-for-1 basis. RUBIA will constitute not more than 1.4% of the total composition of meat and poultry analogue products in which it is used. The use of modified soy protein in meat and poultry analogue products above the specified use-levels is largely self-limiting based on unacceptable organoleptic properties.

As a base case, Impossible Foods has assumed it will capture 1% of the total meat and poultry market with RUBIA containing meat analogue products. 1% of the total meat market represents approximately 5 times the volume of the current meat analogue market size based on sales estimates[4].

The Estimated Daily Intake (EDI) of RUBIA in the target meat analogue applications was established using the Retail Commodity Intakes: Mean Amounts of Retail Commodities per Individual, 2007-08 (Bowman, Martin, Clemens, Lin, & Moshfegh, 2013). The results of that analysis are seen in Table 2. The estimates were calculated as follows:

For beef, the mean daily consumption is 59 grams. Impossible Foods proposes capturing 1% of this market with a beef analogue product consisting of not more than 1.37% RUBIA solids. This equates to an intake of 8.08 mg/person/day of RUBIA. This process was repeated for the other product categories where Impossible Foods will market that contain the modified soy protein RUBIA.

The estimated average daily intake of modified soy protein in the intended applications will be 7.73 mg/person/day and the maximum intake will be 11.81 mg/person/day. As noted above, this base case represents capturing 5 times the existing meat and poultry analogue market.

[4] Datamonitor estimates the US meat analogue volume was 53M kg in 2009. USDA-FAS Livestock and Poultry Report, April 2014 estimates 2014 US consumption of 11B kg beef, 8.5B kg pork, and 14B kg broilers. Therefore, the current meat analogue market is less than 0.2% of the overall meat market and capturing 1% of the meat market represents 5 times the current meat analogue market in the US.

Table 2. Summary of Proposed Uses of RUBIA in Food Applications Based on Retail Food Commodity Intakes 2007-2008.

Food Category to be Replaced	Mean Consumption (gr/day)	Anticipated Market Share Replacement (%)	Anticipated Typical Use rate (%)	Estimated Typical Daily Intake (mg/person/day)	Anticipated Maximum Use Rate (%)	Estimated Maximum Daily Intake (mg/person/day)
Beef	59	1	1.10	6.49	1.37	8.08
Pork	29	1	0.27	0.78	0.68	1.97
Poultry	65	1	0.07	0.46	0.27	1.76
TOTAL				7.73		11.81

Retail Food Commodity Intakes: Mean Amounts of Retail Commodities per Individual, 2007-08. U.S. Department of Agriculture, Agricultural Research Service, Beltsville, MD and US Department of Agriculture, Economic Research Service, Washington, D.C.
http://www.ncaur.usda.gov/SP2UserFiles/Place/12355000/pdf/ficrcd/FICRCD_Intake_Tables_2007_08.pdf

a. Estimation of "90th Percentile" Intake

Following the FDA guidance document "Guidance for Industry: Estimating Dietary Intake of Substances in Food" to estimate daily intake values, the pseudo 90th percentile for modified soy protein consumption would be 2 times the mean EDI or 23.62 mg/person/day.

Therefore, the estimated average daily intake of RUBIA will be between 7.73 mg/person/day and 11.81 mg/person/day at the maximum anticipated use rate. The exposure to high users (90th percentile) will be approximately 23.62 mg/person/day if RUBIA is used at the maximum anticipated rate. As stated above, the estimated mean *per capita* myoglobin consumption is 0.77 g/day and the 90th percentile intake would be 1.54 g/day. Products formulated with anticipated typical rates of RUBIA deliver approximately the same amount of hemeprotein as is found in traditional meat products. Thus, when RUBIA is used at the anticipated typical rate there is no change in the quantity of hemeprotein consumed by meat eaters choosing the meat replacement product.

4.4 Assessment of Allergenicity and Toxicology of Soy Leghemoglobin

Soybeans are acknowledged as a commonly allergenic food. Soybeans are known to contain several allergenic proteins (Taylor, Panda, Goodman, & Baumert, 2014). However, soybean leghemoglobin is not identified among the known soybean allergens. The potential allergenicity of soybean leghemoglobin can be assessed in the same manner as used for the novel proteins expressed in genetically engineered foods. The Codex Alimentarius Commission developed an assessment scheme for the analysis of potential allergenicity of proteins derived from biotechnology (2003). This assessment is a multi-factorial approach which includes assessing the source of the protein for allergenicity, the sequence homology of the protein to known allergens, resistance to pepsin degradation and if there is a high suspicion of allergenicity, specific serum screening. This analysis provides a likelihood of allergenic response by considering the totality of the evidence.

Impossible Foods enlisted Dr. Richard E. Goodman at the Food Allergy Resource and Research Program (FARRP) of the University of Nebraska to assess the potential allergenicity and toxicity of hemoglobin proteins derived from a variety of plants and bacterial sources, consistent with the Codex recommendations. The proteins assessed were derived from soybean (*Glycine max*), barley (*Hordeum vulgare*), mungbean (*Vigna radiata*), a bacterium from the Firmicutes lineage (*Bacillus subtilis*), a bacterium from the Chlamydiae / Verrucomicrobia group (*Methylacidiphilum infernorum*) and a bacterium from Aquificae (*Aquifex aeolicus*). Dr. Goodman also assessed the stability of the soy leghemoglobin protein to pepsin digestion. Dr. Goodman's summary of his findings and copies of the full reports are provided in Annexes 2 thru 4.

Dr. Goodman's assessment included a full literature search to identify any published literature regarding possible allergenicity or toxicity associated with hemoglobin proteins. The literature survey also sought to identify any reports regarding health issues associated with human consumption of hemoglobin proteins of any origin. The conclusion of this assessment was that no published literature could be found that suggested allergic, toxic or adverse health effects related to consumption of any hemoglobin proteins.

The assessment also determined if the amino acid sequence of any of the 6 hemoglobin proteins had sufficient similarity with any known allergen or toxin to suggest possible cross reactivity. The result of this analysis was that none of the target proteins demonstrated sufficient similarity with any known allergen or toxin to raise concern.

Finally, Dr. Goodman assessed the stability of the soy leghemoglobin protein to pepsin degradation. This laboratory-based assessment demonstrated that soy leghemoglobin protein is readily digested by pepsin, confirmed with SDS-PAGE migration and LC-MSMS analysis. Dr. Goodman concluded that consumption of the soy leghemoglobin protein raises no health or safety concern.

4.5 Expert Panel Report

Professor Emeritus
Pharmacology & Toxicology
VCU School of Medicine

JOSEPH F. BORZELLECA

President
Toxicology and Pharmacology, Inc.
Consultative Services

8718 September Drive, Richmond, VA 23229-7319 U.S.A.
Tele: 804.285.2004 Fax: 804.285.1401 Email: josephfborzelleca@comcast.net

The Report of the Expert Panel on the Generally Recognized as Safe Determination of the Proposed Uses of Soybean Leghemoglobin Protein Derived from *Pichia pastoris* as a Food Ingredient

18 August 2014

Introduction

Impossible Foods Inc. (Impossible Foods) convened a panel of independent scientists (Expert Panel), qualified by their scientific training and relevant national and international experience to evaluate the safety of food ingredients, to conduct an independent, critical and comprehensive evaluation of the available safety information on soy leghemoglobin protein, and to determine if the proposed uses as a protein component in meat and poultry replacement (analogue) products of soy leghemoglobin, would be Generally Recognized as Safe (GRAS) based on scientific procedures. The Expert Panel consisted of Professor Joseph F. Borzelleca, Ph.D. (Virginia Commonwealth University School of Medicine), Professor Michael W. Pariza, Ph.D. (University of Wisconsin-Madison), and Professor Steve Taylor (University of Nebraska-Lincoln).

A comprehensive search of the scientific literature on plant and animal hemoglobins and related products through 30 April 2014 was conducted by Impossible Foods. A summary of the results of this search was made available to the Expert Panel. Impossible Foods also provided the Expert Panel with the "GRAS NOTIFICATION FOR SOYBEAN LEGHEMOGLOBIN PROTEIN DERIVED FROM *PICHIA PASTORIS*" (not dated) and a "TECHNICAL SUMMARY OF SOYBEAN LEGHEMOGLOBIN PROTEIN DERIVED FROM *PICHIA PASTORIS* " (dated May 30, 2014). The Expert Panel, independently and collectively, critically evaluated these documents, other information deemed appropriate or necessary and information pertaining to the method of manufacture, product specifications, batch analyses, intended levels of use, exposure estimates, and available scientific information pertaining to the safety of leghemoglobin protein and other plant and animal hemoglobins.

Following independent, critical evaluation of the available information, the Expert Panel convened by teleconference and email correspondence, reviewed its findings with Dr. Don Boudreaux and technical staff of Impossible Foods, and unanimously concluded that the intended uses in meat and poultry products of leghemoglobin protein derived from soybean, manufactured consistent with current Good Manufacturing Practice (cGMP) and meeting appropriate food-grade specifications, are GRAS based on scientific procedures. A summary of the basis for this conclusion appears below.

Impossible Foods proposes to market the symbiotic leghemoglobin isolated from soybean root nodules and produced in the yeast *Pichia pastoris* Bg10 in the United States under the trade name RUBIA for use as a protein component in meat and poultry replacement (analogue) products.

Hemoglobin proteins are found in most organisms, including bacteria, protozoa, fungi, plants and animals (Hardison, 1998). Hemeproteins are classified as globin/non-globin and symbiotic/non-symbiotic. Hemoglobin, myoglobin, and leghemoglobin are examples of globin proteins. Cytochrome oxidases, hemocyanins, and methemalbumin are examples of non-globin hemeproteins (Everse, 2004) (Jokipii-Lukkari, Frey, Kallio, & Haggman, 2009). Plant hemoglobins are classified according to function as symbiotic or non-symbiotic (Gupta, Hebelstrup, Mur, & Igamberdiev, 2011). Symbiotic hemoglobins are found predominantly in leguminous plant species. The most studied symbiotic hemoglobins are the leghemoglobins of nitrogen fixing legumes where they facilitate oxygen diffusion within root tissues. Non-symbiotic hemoglobins have been identified in a wide range of legume and non-legume plants. The highest expression levels for non-symbiotic plant hemoglobin are observed in metabolically active or stressed tissue (Anderson, Jensen, Leewellyn, Dennis, & and Peacock, 1996).

Impossible Foods analyzed sequences of plant hemoglobins, animal myoglobins and plant leghemoglobins including rice, soy, corn, barley, lupine, horse, tuna, and pig species. Structural comparison of plant hemoglobins and animal myoglobins show similarities and differences. Animal myoglobins, plant leghemoglobins and plant hemoglobins adopt the same globin fold and are structurally very similar; all proteins bind a heme prosthetic group involved in binding and/or transport of oxygen. The globin protein family is large, present in a wide range of organisms and is well studied.

Identity and Characterization of Soy Leghemoglobin Protein

The chemical name of the characterizing component of modified soy protein is soy leghemoglobin. The source of the protein is the soybean plant *Glycine max* gene *LGB2*.

There is no Chemical Abstracts Number for this material.

The proposed common or usual name of soy leghemoglobin protein (a protein isolate derived from soy) is "modified soy protein".

Production of Soy Leghemoglobin

The method of production involves four stages: construction of the production strain of *Pichia pastoris* Bg10, expression of soy leghemoglobin protein in submerged fermentation, purification, and stabilization of the expressed soy leghemoglobin protein.

All materials used in the production of soy leghemoglobin are food grade and GRAS or high-quality chemical or pharmaceutical grades (USP, NF, or ACS grades) from approved suppliers and processing conditions are appropriate for food production and consistent with cGMP.

Preparation of the Production Strain for Fermentation

The protein coding sequence from *Glycine max* leghemoglobin *LGB2* was synthesized, codon-optimized for expression in *Pichia pastoris,* and sub-cloned into the commercially available expression vector *pJAN* (BioGrammatics, Inc.) to generate *pJAN-legH.* The plasmid map of *pJAN* is publically available at biogrammatics.com. All components of *pJAN-legH* were confirmed by sequencing. The alcohol oxidase promoter (*pAOX1*) and terminator regulate LegH expression. This promoter has been demonstrated to produce high levels of recombinant proteins after producing biomass on glycerol, and inducing *pAOX1* with methanol. *pJAN-legH* contains two antibiotic resistance genes *AMPR* (beta-lactamase) and *NATR*, which were used only for strain construction purposes Antibiotics were not used during fermentation and RUBIA does not contain the antibiotic resistance genes as demonstrated by quantitative PCR (qPCR) with primers targeting the *AMPR* and *NATR* genes. No DNA was observed at a level of detection of 0.001 pg of DNA (cf. Annex 5, GRASN). Typical yeast transformations require microgram quantities of recombinant DNA for homologous integration; this is approximately 1,000,000,000 times more than the level of detection of this qPCR method. Furthermore, RUBIA does not transform chemically competent *E. coli* cells to acquire AMP or NAT resistance. RUBIA does not contain the production strain MXY022, as demonstrated by the absence of the formation of nourseothricin (NAT) resistant colonies when 0.5 ml of RUBIA is plated on growth medium containing 50 ug/ml NAT and grown for 72 hrs at 30 degrees Celsius (cf. Annex 5, GRASN). On the rare occasion that a NAT-resistant colony is observed, the colony is analyzed microscopically for *Pichia pastoris* morphology and colony PCR is performed to ensure the absence of MXY022 genomic DNA.

P. pastoris Bg10 was transformed with linearized *pJAN-legH* to generate the production strain MXY022. Transformed cells were isolated and characterized by both colony PCR

and quantitative PCR (qPCR) to confirm that a single copy of the *pJAN-LegH* plasmid was integrated into the genome of Bg10 cells at the *AOX1* promoter.

The production strain, *Pichia pastoris* Bg10, was derived from the well characterized strain Y-11430, which is deposited in the collection at the Northern Regional Research Laboratories (NRRL). The lineage of *P. pastoris* strain NRRL Y-11430 was previously included in GRN 204, reviewed by the Agency in 2006.

There are no indications that *P. pastoris* has been associated with animal or human illness. The first *P. pastoris* strains were isolated from an oak tree and a chestnut tree and were deposited in the collection at the Northern Regional Research Laboratories (NRRL) (www.biogrammatics.com). Yeast strains screened by Phillips Petroleum for growth on methanol included two *P. pastoris* strains, designated NRRL Y-1603 (ATCC accession 28485) (ATCC, 2006b) and NRRL YB-4290 (NCAUR, 2006). Phillips Petroleum identified a *P. pastoris* strain with improved growth characteristics. The strain was designated 21-1 and deposited at NRRL, as NRRL Y-11430 . This strain is now available from ATCC as 76273. No records are available confirming that NRRL Y-1603 or NRRL YB-4290 is the progenitor of NRRL Y-11430, but it seems likely that one of them is the progenitor strain. NRRL Y-11430 was the progenitor strain for GS115, a histidine auxotrophic mutant *(his4-)*, a common *Pichia* strain provided in commercial kits by Invitrogen Corporation, and widely used as the parental strain of many biotechnology products, including FDA approved proteins such as Kalibitor® (ecallantide, for the treatment of acute attacks of hereditary angioedema, 2009). Additionally, the GS115 derived strain SMD1168 is used for the GRAS approved production of BD16449 Phospholipase C (Food and Drug Administration, 2006). Like GS115, the BioGrammatics, Inc. strain, Bg10 is also a derivative of NRRL Y-11430, and genomic sequencing data performed by BioGrammatics Inc. confirm the similarity of NRRL Y-11430, Bg10 and GS115. Additional taxonomic history of these strains is available in a 2009 manuscript by C. Kurtzman and on the Biogrammatics webpage (biogrammatics.com).

BioGrammatics, Inc. further developed the NRRL-Y-11430 strain to remove the native *P. pastoris* plasmids using PCR primers unique to the plasmids to screen multiple single-colony isolates for the presence of the plasmids. One isolate without plasmids was selected to become the wild-type (wt) BioGrammatics strain, Bg10. Genomic sequence from Bg10 indicates the plasmids are no longer present, and, benchmarks the similarity of Bg10 with NRRL-Y11430, as well as with GS115. Like NRRL Y-11430 and GS115, Bg10 does not contain antibiotic resistance genes.

Expression of Soy Leghemoglobin Protein in Submerged Fermentation, Purification and Stabilization

Soy leghemoglobin protein is obtained by fed-batch fermentation using *P. pastoris* Bg10. All media components are FCC approved or food-grade ingredients. The *P. pastoris* cells in the fermentation broth are concentrated (separated) via centrifugation and then lysed by high-pressure homogenization. Insoluble material within the lysate is removed by filtration. Soy leghemoglobin protein is purified from the clarified lysate by chromatographic separation. Ultrafiltration is used to concentrate soy leghemoglobin protein up to 80 g/l. The resulting concentrated sample is 0.2 um filtered and stored either as a frozen liquid or a dehydrated powder.

Specifications for Soy Leghemoglobin

Following purification, RUBIA is standardized to contain 80 grams per liter (g/l) soy leghemoglobin protein. Potassium phosphate and sodium chloride are used to stabilize the product. All stabilizing agents are food grade. The product specifications are presented below. The mean % composition from three batches of liquid RUBIA is presented to demonstrate consistency of production.

	RUBIA Specifications		Batch Analysis (Frozen Liquid) (% w/w)		
	Concentration (Frozen Liquid) (% w/w)	Concentration (Dry Powder) (% w/w)	PP-PGM2-14-120	PP-PGM2-14-125	PP-PGM2-14-127
Protein	10	91	9.71	9.76	10.03
Leghemoglobin	8	73	7.90	8.09	8.02
Ash	<1	<9	0.84	0.84	0.85
Fat	<0.1	<1	0.03	0.07	0.07
Carbohydrate	<0.1	<1	<0.1	0.11	<0.1
Solids	11	100	10.24	10.78	10.51
Moisture	89	0	89.76	89.22	89.49

	RUBIA Specifications		Batch Analysis (Frozen Liquid) (ppm w/w)		
	Concentration (Frozen Liquid) (ppm w/w)	Concentration (Dry Powder) (ppm w/w)	PP-PGM2-14-120	PP-PGM2-14-125	PP-PGM2-14-127
Lead	<0.01	<0.1	<0.01	<0.01	<0.01
Arsenic	<0.01	<0.1	<0.01	<0.01	<0.01
Mercury	<0.005	<0.05	<0.005	<0.005	<0.005
Cadmium	<0.1	<1	<0.1	<0.1	<0.1

	RUBIA Specifications		Batch Analysis (Frozen Liquid)		
	Concentration (Frozen Liquid)	Concentration (Dry Powder)	PP-PGM2-14-120	PP-PGM2-14-125	PP-PGM2-14-127
Aerobic plate count[1] (CFU/g)	<10^4	<10^4	<10	<10	<10
E. coli O157H7[2]	Absent by test	Absent by test	Absent by test	Absent by test	Absent by test
Salmonella spp[3]	Absent by test	Absent by test	Absent by test	Absent by test	Absent by test
Listeria monocytogenes[4]	Absent by test	Absent by test	Absent by test	Absent by test	Absent by test

[1] AOAC OMA 990.12
[2] AOAC RI 020801
[3] AOAC OMA 2011.03
[4] AOAC OMA 2010.02

Stability of Soy Leghemoglobin

The soy leghemoglobin product (RUBIA) may be stored at -20 °C as a frozen liquid or dried powder for a minimum of 12 months with no observable change in leghemoglobin stability or performance in meat and poultry analogue products.

Intended Uses in Food

Soy leghemoglobin is proposed to be used as a plant-based protein component in non-animal derived food products with the texture, nutrition, flavor and appearance of traditional animal derived foods. Soy leghemoglobin will impart a unique flavor impact to meat analogue products. The high bioavailability of the heme iron component of soy leghemoglobin (modified soy protein) makes it suitable to enhance the dietary profile of many processed foods (Carpenter & Mahoney, 1992).

Soy leghemoglobin may be one of several plant proteins approved by FDA that will comprise meat and poultry analogue products. Other proteins may include, but are not limited to, commercially available proteins from soy, pea, mung bean, lentil, corn, and wheat. Soy leghemoglobin will function to contribute to the flavor, appearance and nutritional quality of meat analogue products. A typical meat analogue product may contain:

Component	Meat Analogue
Protein	10%-25%
Plant Oils	0%-25%
Miscellaneous+	2%
Water	60%-75%

+Miscellaneous ingredients may include salt, flavors, spices and seasonings, emulsifiers.

RUBIA will be added to the meat or poultry analogue product to deliver not more than 1% modified soy protein. RUBIA is formulated to contain 73% modified soy protein on a solids basis. Thus, RUBIA will constitute not more than 1.4% of the total composition of meat and poultry analogue products in which it is used.

Self-limitation of the Use of Soy Leghemoglobin

The use of soy leghemoglobin (modified soy protein) in meat and poultry analogue products above the specified use-levels is largely self-limiting based on unacceptable organoleptic properties.

Estimated Dietary Intake

The vast majority of hemeproteins consumed in the diet are as myoglobin through consumption of meat and poultry products. Hemeprotein consumption was estimated using data from the "Retail Commodity Intakes: Mean Amounts of Retail Commodities per Individual, 2007-08. (Bowman, Martin, Clemens, Lin, & Moshfegh, 2013). For the US population, per capita mean consumption of meat and poultry products is 154 g/day and the 90th percentile intake is 308 g/day. Assuming an average myoglobin concentration for meat and poultry products of 0.5% (Yip & Dallman, 1996), the average per capita myoglobin consumption would be 0.77 g/day and the 90th percentile intake would be 1.54 g/day.

Soy leghemoglobin will be marketed for use in meat and poultry analogue products that provide consumers a flavorful and nutritious alternative to meat containing products. Impossible Foods has estimated daily intakes of soy leghemoglobin by assuming

consumers will substitute the meat analogue product for the traditional meat product on a 1-for-1 basis.

As a base case, Impossible Foods has assumed it will capture 1% of the total meat and poultry market with RUBIA containing meat analogue products. 1% of the total meat market represents approximately 5 times the volume of the current meat analogue market size based on sales estimates[1]. The Estimated Daily Intake (EDI) of RUBIA in the target meat analogue applications was established using the Retail Commodity Intakes: Mean Amounts of Retail Commodities per Individual, 2007-08 (Bowman, Martin, Clemens, Lin, & Moshfegh, 2013). The results of that analysis are presented below. The estimates were calculated as follows. For beef, the mean daily consumption is 59 grams. Impossible Foods proposes capturing 1% of this market with a beef analogue product consisting of not more than 1.37% RUBIA solids. This equates to an intake of 8.08 mg/person/day of RUBIA. This process was repeated for the other product categories where Impossible Foods will market that contain the modified soy protein RUBIA.

The estimated average daily intake of modified soy protein in the intended applications will be 7.73 mg/person/day and the maximum intake will be 11.81 mg/person/day. As noted above, this base case represents capturing 5 times the existing meat and poultry analogue market.

Summary of Proposed Uses of Soy leghemoglobin in Food Applications Based on Retail Food Commodity Intakes 2007-2008.

Food Category to be Replaced	Mean Consumption (gr/day)	Anticipated Market Share Replacement (%)	Anticipated Typical Use rate (%)	Estimated Typical Daily Intake (mg/person/day)	Anticipated Maximum Use Rate (%)	Estimated Maximum Daily Intake (mg/person/day)
Beef	59	1	1.10	6.49	1.37	8.08
Pork	29	1	0.27	0.78	0.68	1.97

[1] Datamonitor estimates the US meat analogue volume was 53M kg in 2009. USDA-FAS Livestock and Poultry Report, April 2014 estimates 2014 US consumption of 11B kg beef, 8.5B kg pork, and 14B kg broilers. Therefore, the current meat analogue market is less than 0.2% of the overall meat market and capturing 1% of the meat market represents 5 times the current meat analogue market in the US.

Poultry	65	1	0.07	0.46	0.27	1.76
TOTAL				7.73		11.81

Retail Food Commodity Intakes: Mean Amounts of Retail Commodities per Individual, 2007-08. U.S. Department of Agriculture, Agricultural Research Service, Beltsville, MD and US Department of Agriculture, Economic Research Service, Washington, D.C.
http://www.ncaur.usda.gov/SP2UserFiles/Place/12355000/pdf/ficrcd/FICRCD_Intake_Tables_2007_08.pdf

Safety of Soy Leghemoglobin

The safety and the suitability of the proposed uses of soy leghemoglobin are based on the chemical and functional equivalency of soy leghemoglobin to animal myoglobins, plant hemoglobins and plant leghemoglobins.

Chemical and Functional Equivalency

Hemeproteins are found in most organisms, including bacteria, protozoa, fungi, plants and animals (Everse, 2004) (Hardison, 1998). Soybeans have been shown to express three hemoglobin proteins: symbiotic, non-symbiotic and truncated (Lee, Kim, & An, 2004). Symbiotic plant hemoglobins, which evolved from non-symbiotic hemoglobins (Gupta, Hebelstrup, Mur, & Igamberdiev, 2011), are commonly referred to as leghemoglobins. Symbiotic leghemoglobins, found predominately in legume root nodules, function in the nitrogen fixation process in concert with the bacterium *Rhizobium* where they facilitate oxygen diffusion within host root tissues. RUBIA contains this symbiotic leghemoglobin derived from soybean.

Anderson et al. demonstrated that the non-symbiotic hemoglobin in soybeans was expressed in various plant tissues including stems, shoots, cotyledon, leaves, and root hair (Anderson, Jensen, Leewellyn, Dennis, & and Peacock, 1996). These soybean tissues are commonly consumed in the diet in the form of bean sprouts. Commercial production of soybean sprouts is a 6 day process from imbibition to packaging for retail sale (Lim, 2014). Sprouted barley, which is widely used in the beverage industry (malted barley) and in the baking industry (malted barley flour), has been shown to express hemoglobin 1 day after imbibition (Duff, Guy, Xianzhou, Durnin, & Hill, 1998). Non-symbiotic hemoglobins are expressed in the rice embryo as well as in the coleoptiles and seminal root of sprouted rice, which is consumed as part of the diet (Lira-Ruan, Ruiz-Kubli, & Arredondo-Peter, 2011).

Impossible Foods analyzed plant symbiotic leghemoglobins (soy, lupine), non-symbiotic plant hemoglobins (rice, corn, barley), and animal myoglobins (horse, tuna, pig) and confirmed the structural similarity (cf. Annex 1, GRASN). The three dimensional

structure of soybean leghemoglobin is highly similar to the non-symbiotic hemoglobins of corn, rice, and barley as well as mammalian myoglobin.

All globin proteins bind the heme prosthetic group and are involved in binding or transporting oxygen. The oxygen binding mechanism of soy leghemoglobin is similar to that of animal muscle myoglobin.

History of Safe Use

The safety of soy protein is well established. Soybeans have been part of the human diet for more than 5000 years.

In the 2004/2005 marketing year, 229 million metric tons of soybeans were produced worldwide. Although the majority of the crop is used for animal feed, approximately 14% is used for human food in the form of traditional soyfoods, e.g. tofu, soymilk, natto, miso, bean sprouts, and as soy protein ingredients used to formulate food products as diverse as infant formula, dairy and meat alternatives, nutritional supplements and energy bars. (Golbitz & Jordan, 2006) Plant and animal hemoglobin proteins are widely consumed in the human diet where they represent a highly bioavailable source of dietary iron for human nutrition. Plant-derived hemoglobins are prevalent in our food system through malted grain products and sprouted beans (pulses).

Regulatory Status

The use of soy proteins is widely accepted in the United States. The US Food and Drug Administration has affirmed the safety of soy protein isolates for inclusion in many products and has approved a health claim for soy protein and the reduced risk of coronary heart disease (21 CFR 101.82). In 2000, the US Department of Agriculture issued a ruling allowing soy protein to completely replace animal protein in the National School Lunch Program (Messina, 2006). The safety of soy in human food has been clearly established and affirmed by the two major food regulatory agencies in the US.

Allergenicity

Soybeans are acknowledged as a commonly allergenic food. Soybeans are known to contain several allergenic proteins (Taylor, Panda, Goodman, & Baumert, 2014). However, soybean leghemoglobin is not identified among the known soybean allergens. The potential allergenicity of soybean leghemoglobin can be assessed in the same manner as used for the novel proteins expressed in genetically engineered foods. The Codex Alimentarius Commission developed an assessment scheme for the analysis of potential allergenicity of proteins derived from biotechnology (Codex Alimentarius, 2003). This assessment is a multi-factorial approach which includes assessing the source of the protein for allergenicity, the sequence homology of the protein to known

allergens, resistance to pepsin degradation and, if there is a high suspicion of allergenicity, specific serum screening. This analysis provides a likelihood of allergenic response by considering the totality of the evidence.

Impossible Foods enlisted Dr. Richard E. Goodman at the Food Allergy Resource and Research Program (FARRP) of the University of Nebraska to assess the potential allergenicity and toxicity of hemoglobin proteins derived from a variety of plants and bacterial sources, consistent with the Codex recommendations. The proteins assessed were derived from soybean (*Glycine max*), barley (*Hordeum vulgare*), mungbean (*Vigna radiata*), a bacterium from the Firmicutes lineage (*Bacillus subtilis*), a bacterium from the Chlamydiae / Verrucomicrobia group (*Methylacidiphilum infernorum*) and a bacterium from Aquificae (*Aquifex aeolicus*). Dr. Goodman conducted a comprehensive search of the biomedical literature to identify any published reports regarding possible allergenicity or toxicity associated with hemoglobin proteins and any reports regarding health issues associated with human consumption of hemoglobin proteins of any origin. Dr. Goodman concluded "My conclusion from this "weight of evidence" approach to dietary protein safety is that the soybean leghemoglobin is very unlikely to present a risk of dietary allergy or toxicity to consumers."
Bioinformatics searches (amino acid sequence comparisons) were performed comparing the known sequence of soybean leghemoglobin (GI:126241) with known or putative allergens listed in the AllergenOnline.org, version 13 database using both FASTA full-length sequence alignments and search for 80 amino acid matches along the entire sequence looking for >35% identity. No significant alignments were found (>50% identity over most of the length of the leghemoglobin protein). A search using the conservative sliding window of 80 amino acid algorithm did reveal a 35.2% identity match with potato patatin, a minor allergen. However, there was only a single segment alignment (35.2% identity) in a highly gapped overlap of 105 amino acids to potato patatin. The gapping and high lack of identified alignment to this protein in full-length FASTA alignment demonstrate that the match was a random match without scientific basis for concern that soy leghemoglobin and potato patatin would cross-react. In addition, no published reports exist of allergic cross-reactivity between potato and soybean. Thus, the bioinformatics searches did not reveal any similarities of concern between soybean leghemoglobin and known allergens.

Similarly, bioinformatics searches were conducted to determine if similarities existed between the amino acid sequence of soybean leghemoglobin and the sequences of known toxic proteins. A BLASTP search of *Glycine max* leghemoglobin (GI:126241) with the NCBI protein database using the keyword limit "toxin" did not uncover any significant alignments with known toxic proteins. These results do not raise concerns of potential toxicity for the soybean leghemoglobin protein.

Additionally, Dr. Goodman's laboratory tested the stability of soybean leghemoglobin in a model simulated gastric digestion study using the conditions recommended by Ofori-Anti et al. (Ofori-Anti, Ariyarathna, Chen, Lee, Pramod, & Goodman, 2008) to evaluate the pepsin stability of novel proteins in genetically modified crops. A positive correlation exists between the stability of abundant dietary proteins in this assay and the likelihood that they will be identified as food allergens. The soybean leghemoglobin was very rapidly digested by pepsin (90% with 0.5 min and no detectable residue after 1 min). No stable fragments of the leghemoglobin protein were detected either. On the basis of resistance to pepsin digestion, soybean leghemoglobin shows a low potential risk of allergenicity or toxicity.

Dr. Goodman stated "My conclusion from this "weight of evidence" approach to dietary protein safety is that the soybean leghemoglobin is very unlikely to present a risk of dietary allergy or toxicity to consumers"

Conclusions

We, members of the Expert Panel, have individually and collectively critically evaluated the information and data summarized above and other information deemed pertinent to the safety of the proposed uses of soy leghemoglobin (a symbiotic hemoglobin protein derived from soybean root nodules by Impossible Foods Inc.). We unanimously conclude that the proposed uses as a protein component in meat and poultry replacement (analogue) products of soy leghemoglobin, produced consistent with current Good Manufacturing Practice (cGMP) and meeting the food grade specifications presented above, are safe and suitable.

We unanimously conclude that the proposed uses as a protein component in meat and poultry replacement (analogue) products of soy leghemoglobin, produced consistent with current Good Manufacturing Practice (cGMP) and meeting the food grade specifications presented above are Generally Recognized As Safe (GRAS) based on scientific procedures (structural and functional equivalency to other safe plant and animal hemoglobins).

It is our unanimous opinion that other qualified experts would concur with these conclusions.

(b) (6)

Professor Joseph F. Borzelleca, Ph.D.
Virginia Commonwealth University School of Medicine

19 August 2014
Date

(b) (6)

Professor Michael W. Pariza, Ph.D.
University of Wisconsin-Madison

27 Aug 2014
Date

(b) (6)

Professor Steve Taylor
University of Nebraska- Lincoln

22 August 2014
Date

References

Anderson, C. R., Jensen, E. O., Leewellyn, D. J., Dennis, E. S., & and Peacock, W. J. (1996, June). A new hemoglobin gene from soybean: A role for hemoglobin in all plants. *Proc. Natl. Acad. Sci. USA, 93*, 5682-5687.

Bowman, S., Martin, C., Clemens, J., Lin, B.-H., & Moshfegh, A. (2013). *Retail Commodity Intakes: Mean Amounts of Retail Commodities per Individual, 200702008.* Agricultural Research Service, Food Surveys Research Group. Washington, D.C.: U.S. Department of Agriculture.

Carpenter, C. E., & Mahoney, A. (1992). Contributions of heme and nonheme iron to human nutrition. *Critical Reviews in Food Science and Nutrition*, 333-367.

Codex Alimentarius. (2003). *AppendixIII, 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.* Rome: Codex Alinorm 03/34: Joint FAO/WHO Food Standard Programme, Codex Alimentarius Commission, Twenty-Fifth Session, Rome, Italy 30 June-5 Jult, 2003. .

Duff, S. M., Guy, P. A., Xianzhou, N., Durnin, D. C., & Hill, R. D. (1998). Haemoglobin expression in germinating barley. *Seed Science Research*, 431-436.

Everse, J. (2004). Heme Proteins. In *Encycloedia of Biollogical Chemistry, Volume 2* (pp. 354-361). Elseview Inc.

Food and Drug Administration. (2006). *GRN 204: Phospholipase C enzyme preparation from Pichia pastoris expressing a heterologous phospholipase C gene.* Washington, D.C.

Golbitz, P., & Jordan, J. (2006). Soyfoods: Market and Products. In N. N. Riaz, *Soy Applications in Food* (pp. 1-22). Boca Raton, FL: CRC Press.

Gupta, K. J., Hebelstrup, K. H., Mur, L. A., & Igamberdiev, A. U. (2011). Plant hemoglobins: Important players at the crossroads between oxygen and nitric oxide. *FEBS Letters*, 3843-3849.

Hardison, R. (1998). Hemoglobins from bacteria to man: Evolution of different paterns of gene expression. *The Journal of Experimental Biology*, 1099-1117.

Jokipii-Lukkari, S., Frey, A. D., Kallio, P. T., & Haggman, H. (2009). Intrinsic non-symbiotic and truncated haemoglobins and heterologous Vitreocilla haemoglobin expression in plants. *Journal of Experimental Botany*, 409-422.

Lee, H., Kim, H., & An, C. S. (2004). Cloning and Expression Analysis of 2-on-2 Hemoglobin from Soybean. *Journal of Plant Biology, 47*(2), 92-98.

Lim, O. (2014, April 21). Manager, Food One, Commercial Producer of Soybean Sprouts, Lilburn, GA. (D. P. Boudreaux, Interviewer)

Lira-Ruan, V., Ruiz-Kubli, M., & Arredondo-Peter, R. (2011). Expression of non-symbiotic hemoglobin 1 and 2 genes in rice (Oryza sativa) embryonic organs. *Communicative and Integrative Biology, 4*(4), 457-458.

Messina, M. (2006). Overview of the Health Effects of Soyfoods. In N. N. Riaz, *Soy Applications in Food* (pp. 23-38). Boca Raton, FL: CRC Press.

NCAUR. (2006). Retrieved June 28, 2014, from National Center for Agrticultural Utilization Research: http://nrrl.ncaur.usda.gov/cgi-bin/usda/yeast/report.html?nrrlcodes=YB-4290

Ofori-Anti, A., Ariyarathna, H., Chen, L., Lee, H., Pramod, S., & Goodman, R. (2008). Establishing objective detection limits for the pepsin digestion assay used in the assessment of genetically modified foods. *Reg Toxicol Pharmacol*, 52:94-103.

Olempska-Beer, Z., Merker, R., Ditto, M., & DiNovi, M. (2006). Food-processing enzymes from recombinant microorganisms- A Review. *Regulatory Toxicology and Pharmacology, 45*, 144-158.

Pariza, M., & Johnson, E. (2001). Evaluating the safety of microbial enzyme preparations used in food processing: update for a new century. *Regulatory Toxicology and Pharmacology, 33*, 173-186.

Taylor, S., Panda, R., Goodman, R., & Baumert, J. (2014). Soybeans as a food allergen. In S. Flanagan (Ed.), *Handbook of Food Allergen Detection and Control* (p. In Press). Cambridge, UK: Wookhead Publishing Ltd.

Yip, R., & Dallman, P. R. (1996). IRON. In E. E. Ziegler (Ed.), *Present Knowledge in Nutrition* (Seventh ed., pp. 277-292). Washington, DC: ILSI Press.

4.6 GRAS Conclusion

Impossible Foods Inc. (Impossible Foods) convened a panel of independent scientists (Expert Panel), qualified by their scientific training and relevant national and international experience to evaluate the safety of food ingredients, to conduct an independent, critical and comprehensive evaluation of the available safety information on soy leghemoglobin protein, and to determine if the proposed uses as a protein component in meat and poultry replacement (analogue) products of soy leghemoglobin, would be Generally Recognized as Safe (GRAS) based on scientific procedures.

The Expert Panel, independently and collectively, critically evaluated the documents provided by Impossible Foods and other information deemed appropriate or necessary and information pertaining to the method of manufacture, product specifications, batch analyses, intended levels of use, exposure estimates, and available scientific information pertaining to the safety of leghemoglobin protein and other plant and animal hemoglobins.

The panel reviewed reports prepared by Dr. Richard E. Goodman (Food Allergy Resource and Research Program (FARRP) of the University of Nebraska) that assessed of the potential allergenicity and toxicity of hemoglobin proteins derived from a variety of plants and bacterial sources, consistent with the Codex recommendations. Dr. Goodman conducted a comprehensive search of the biomedical literature to identify any published reports regarding possible allergenicity or toxicity associated with hemoglobin proteins and any reports regarding health issues associated with human consumption of hemoglobin proteins of any origin. Dr. Goodman concluded that there is no published evidence that soybean leghemoglobin or other similar proteins were associated with allergic reactions or toxicity.

Dr. Goodman conducted bioinformatics searches (amino acid sequence comparisons) that compared the known sequence of soybean leghemoglobin (GI:126241) with known or putative allergens. The bioinformatics searches did not reveal any similarities of concern between soybean leghemoglobin and known allergens. Bioinformatics searches were also conducted to determine if similarities existed between the amino acid sequence of soybean leghemoglobin and the sequences of known toxic proteins. These results did not raise concerns of potential toxicity for the soybean leghemoglobin protein.

Additionally, Dr. Goodman's laboratory tested the stability of soybean leghemoglobin in a model simulated gastric digestion study using the conditions recommended by Ofori-Anti et al. (Ofori-Anti, Ariyarathna, Chen, Lee, Pramod, & Goodman, 2008) to evaluate the pepsin stability of novel proteins in genetically modified crops. The soybean leghemoglobin was very rapidly digested by pepsin. On the basis of resistance to pepsin digestion, soybean leghemoglobin shows a low potential risk of allergenicity or toxicity.

Taken together, this weight of evidence approach to the safety assessment of soybean leghemoglobin demonstrates that the protein is very unlikely to present a dietary safety risk to consumers.

The Expert Panel unanimously concluded that the intended uses in meat and poultry replacement (analogue) products of leghemoglobin protein derived from soybean, manufactured consistent with current Good Manufacturing Practice (cGMP) and meeting appropriate food-grade specifications, are GRAS based on scientific procedures.

LIST OF REFERENCES

AAFCO. (2013). 2013 Official Publication. USA: American Association of Feed Control Officials Inc.

Anderson, C. R., Jensen, E. O., Leewellyn, D. J., Dennis, E. S., & and Peacock, W. J. (1996, June). A new hemoglobin gene from soybean: A role for hemoglobin in all plants. *Proc. Natl. Acad. Sci. USA, 93*, 5682-5687.

Aunstrup, K., Andersen, O., Falch, E. A., & Nielsen, T. K. (1979). Production of Microbial Enzymes. In *Microbial Technology* (2 ed ed., pp. 281-309). Academic Press.

Berkowitz, D., & Maryanski, J. (1989). Implications of biotechnology on internatinal food standards and codes of practice. In *FAO/WHO Food Standards Programme.* Geneva: Codex Alimentarius Commission Eighteenth Session.

Bodnar, A. L. (2011). Novel approaches to improving qualities of maize endosperm. *Graduate Theses and Dissertations. Paper 10271*. Ames, Iowa, USA: Iowa State University Digital Repository @ Iowa State University.

Bowman, S., Martin, C., Clemens, J., Lin, B.-H., & Moshfegh, A. (2013). *Retail Commodity Intakes: Mean Amounts of Retail Commodities per Individual, 200702008.* Agricultural Research Service, Food Surveys Research Group. Washington, D.C.: U.S. Department of Agriculture.

Carpenter, C. E., & Mahoney, A. (1992). Contributions of heme and nonheme iron to human nutrition. *Critical Reviews in Food Science and Nutrition*, 333-367.

Center for Disease Control. (1999). Biosafety in Microbiological and biomedical Laboratories. *CDC Office of Health and Safety*, 17-53.

Cereghino, L. J., & Cregg, J. M. (2000). Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiol Rev, 24*, 45–66.

Codex Alimentarius. (2003). *AppendixIII, 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.* Rome: Codex Alinorm 03/34: Joint FAO/WHO Food Standard Programme, Codex Alimentarius Commission, Twenty-Fifth Session, Rome, Italy 30 June-5 Jult, 2003. .

Cregg, J., Barringer, K., Hessler, A., & Madden, K. (1985). Pichia pastoris as a host system for transformations. *Mol.Cel1 Biol. 5,3376-3385, 5*, 3376-3385.

Duff, S. M., Guy, P. A., Xianzhou, N., Durnin, D. C., & Hill, R. D. (1998). Haemoglobin expression in germinating barley. *Seed Science Research*, 431-436.

Enzyme Technical Association. (2005). Enzyme technical association position on food allergen labeling of microbial derived enzymes under FALCPA as it applies to fermentation media raw materials. 1-2.

EU Scientific Committee for Food. (1992). *Report for the Scientific Committee for Food (Twenty-seventy series) 13-22.* Commission of the European Communitties.

Everse, J. (2004). Heme Proteins. In *Encycloedia of Biollogical Chemistry, Volume 2* (pp. 354-361). Elseview Inc.

FAO/WHO. (1996). *Biotechnology and food safety. Report of a Joint FAO/WHO Consultation. FAO Food and Nutrition Paper 61.*

FDA 21 CFR Part 573. (1993). [Docket No. 87F-02211 Food additives permitted in feed and drinking water of animals: Pichia Pastoris dried yeast. *Federal Register, 58*, 59169-59170.

Food and Agricultural Organization of the United Nations. (2010). *Food and Agricultural Commodity Production*. Retrieved May 13, 2014, from http://faostat.fao.org/

Food and Drug Administration. (2004, Jan 8). *Letter to Kyowa Hakko USA, Inc. re: GRAS Notice No. GRN 134 datedd Jan. 8, 2004*.

Food and Drug Administration. (2006). *GRN 204: Phospholipase C enzyme preparation from Pichia pastoris expressing a heterologous phospholipase C gene.* Washington, D.C.

Food and Drug Administration. (2006, July 20). Letter to Sidley, Austin, Brown and Wood, LLP. re: GRAS Notice No. GRN 186 datedd July 20, 2006.

Food and Drug Administration. (2009, Sept. 4). Letter to Monsanto Company re: GRAS Notice No. GRN 283 datedd Sept. 4, 2009.

Food and Drug Administration. (2010, April 7). Letter to Monsanto Company re: GRAS Notice No. GRN 306 datedd Apr. 70, 2010.

Golbitz, P., & Jordan, J. (2006). Soyfoods: Market and Products. In N. N. Riaz, *Soy Applications in Food* (pp. 1-22). Boca Raton, FL: CRC Press.

Gupta, K. J., Hebelstrup, K. H., Mur, L. A., & Igamberdiev, A. U. (2011). Plant hemoglobins: Important players at the crossroads between oxygen and nitric oxide. *FEBS Letters*, 3843-3849.

Hardison, R. (1998). Hemoglobins from bacteria to man: Evolution of different paterns of gene expression. *The Journal of Experimental Biology*, 1099-1117.

Hargrove, M. S., Barry, J., Brucker, E. A., Berry, M. B., Phillips, G. N., Olson, J. S., et al. (1997). Characterization of recombinant soybean leghemoglobin a and apolar distal histidine mutants. *Journal of molecular biology*, 1032-42.

Herbst, R., Hess, K., Tran, H., Tseng, J., Mullani, N., Charnsangavej, C., et al. (2002). Phase I study of recombinant human endostatin in patients with advanced solid tumors. *J.Clin.Oncol, 20*, 3792-3803.

International Food Biotechnology Council. (1990). Chapter 4: Safety evaluation of foods and food ingredients derived from microorganisms. *Regul Toxicol Pharmacol*, S114-S128.

Jokipii-Lukkari, S., Frey, A. D., Kallio, P. T., & Haggman, H. (2009). Intrinsic non-symbiotic and truncated haemoglobins and heterologous Vitreocilla haemoglobin expression in plants. *Journal of Experimental Botany*, 409-422.

Jonas, D., Antignac, E., Antoine, J., Classen, H., Huggett, A., Knudsen, I., et al. (1996). The safety assessment of novel foods. Guidlines prepared by LSI European Novel Food Task Force. *Food Chem Tocicol, 34*, 931-940.

Kurtzman, C. P. (2009). Biotechnological strains of Komagataella (Pichia) pastoris are Komagataella phaffii as determined from multigene sequence analysis. *Society for Industrial Microbiology*.

Lee, G.-A., Crawford, G. W., Liu, L., Sasaki, Y., & Chen, X. (2011). Archaeological Soybean (Glycine max) in East Asia: Does Size Matter? *PLoS One, 6*, 11.

Lee, H., Kim, H., & An, C. S. (2004). Cloning and Expression Analysis of 2-on-2 Hemoglobin from Soybean. *Journal of Plant Biology, 47*(2), 92-98.

Lira-Ruan, V., Ruiz-Kubli, M., & Arredondo-Peter, R. (2011). Expression of non-symbiotic hemoglobin 1 and 2 genes in rice (Oryza sativa) embryonic organs. *Communicative and Integrative Biology, 4*(4), 457-458.

Madden, K.R. (2014, May). *Ph. D.* BioGrammatics, Inc. Citation approval for strain lineage. Written Communication to Impossible Foods.

Messina, M. (2006). Overview of the Health Effects of Soyfoods. In N. N. Riaz, *Soy Applications in Food* (pp. 23-38). Boca Raton, FL: CRC Press.

OECD. (1992). *Safety Considerations for Biotechnology.* Retrieved 2014, from http://dbtbiosafety.nic.in/guideline/OACD/Safety_Considerations_for_Biotechnology_1992.pdf

OECD. (1993). *OEDC safety Evaluation of Foods Derived by Modern Biotechnology Concepts and Principles.* Retrieved from www.oecd.org/dataoecd/57/3/194629.pdf

Olempska-Beer, Z., Merker, R., Ditto, M., & DiNovi, M. (2006). Food-processing enzymes from recombinant microorganisms- A Review. *Regulatory Toxicology and Pharmacology, 45*, 144-158.

Pariza, M., & Foster, E. (1983). Determining the safety of enzymes used in food processing. *J. Food Prot.*, 453-468.

Pariza, M., & Johnson, E. (2001). Evaluating the safety of microbial enzyme preparations used in food processing: update for a new century. *Regulatory Toxicology and Pharmacology, 33*, 173-186.

Proulx, A. K., & Reddy, M. B. (2006). Iron bioavailability of hemoglobin from soy root nodules using a Caco-2 cell culture model. *J Agric Food Chem*, 1518-1522.

Riaz, M.N. (2006). *Soy applications in food.* Boca Raton, FL: CRC Press, Taylor & Francis Group.

Rij, N. J., & Kreger-van, W. (1984). *The Yeasts — a Taxonomic Study.* NY, NY: Elsevier Science Publishers.

Schanz, J. (n.d.). *Uses for Peanut Plant Roots.* Retrieved January 27, 2014, from www.ehow.com/list_6580849_yses-peanut-plant-roots.html: www.ehow.com/list_6580849_yses-peanut-plant-roots.html

Taylor , S. L., & Baumert, J. L. (2013). Testing Microbially Derived Enzymes for Potential Allergens From Fermentation Media Raw Materials. *Expert Opinion Statement: Food Allergy Research and Resource Program*, 1-2.

Taylor, S., Panda, R., Goodman, R., & Baumert, J. (2014). Soybeans as a food allergen. In S. Flanagan (Ed.), *Handbook of Food Allergen Detection and Control* (p. In Press). Cambridge, UK: Wookhead Publishing Ltd.

U.S. Department of Agriculture, A. R. (1997). *Results from USDA's 1994-1996 Continuing Survey of Food Intakes by Individuals and 1994-1996 Diet and Health Knowledge Survey.* Washington, D.C.: U.S.D.A., Agricultural Research Service.

Wajcman, H., & Kiger, L. (2002). Hemoglobin, from microorganism to man: a single structural motif, multiple functions. *Comptes Rendus Biologies, 325*(12), 1159-1174.

Wegner, E. (2006, Oct 14). *Patent No. 4,617,274.* USA.

Yip, R., & Dallman, P. R. (1996). IRON. In E. E. Ziegler (Ed.), *Present Knowledge in Nutrition* (Seventh ed., pp. 277-292). Washington, DC: ILSI Press.

LIST OF ANNEXES

Annex 1. Structural comparison of plant hemoglobins and animal myoglobins

Annex 2. Dr. Richard E. Goodman Expert Opinion of Assessment of the Safety of Soy Leghemoglobin

Annex 3. Bioinformatic analysis of six heme proteins from diverse taxonomic sources for potential use in foods

Annex 4. Leghemoglobin *in vitro* digestibility study in human simulated gastric fluid

Annex 5. Batch analyses for three independent lots of RUBIA

Annex 6. Detection of non-symbiotic hemoglobin protein in mung bean sprouts

Annex 1
Structural comparison of plant hemoglobins and animal myoglobins

The globin structural superfamily is a large, well studied family of globular proteins, present in all domains of life: archae, bacteria, and eukaryotes (PFAM PF00042). All members of the globin structural superfamily are thought to share a common ancestor (Punta et al. 2012). The globin structural fold is comprised of eight alpha helical segments and a heme co-factor, which coordinates binding and/or transfer of oxygen. Structural comparisons of animal myoglobin, plant leghemoglobin, and plant non-symbiotic hemoglobin monomers are shown in Figure 1A-H. The crystal structure for cow myoglobin does not exist, so we have included myoglobin structures from tuna, pig, and horse in this analysis. Based on their similarity to eachother (Figure 1F-H), we expect that they are highly similar to cow myoglobin. The crystal structures were superimposed over all backbone atoms using the Super algorithm in PyMOL (Delano, 2007) (Figure 1I-L) and the corresponding root mean square deviations (RMSDs) are shown in Table 1 . Comparison of proteins folds (Figure 1) and RMSD values (Table 1) illustrates that animal myoglobins, plant non-symbiotic hemoglobins, and plant leghemoglobins all adopt the same globin fold and are structurally very similar. Furthermore, animal myoglobins, plant non-symbiotic hemoglobins, and plant leghemoglobins all bind the identical heme prosthetic group, heme B (Figure 1M).

Leghemoglobins, non-symbiotic hemoglobins, and myoglobins each contain the identical heme b co-factor (Figure 1M). Soybean leghemoglobin does not contain peptide sequences that are associated with allergenicity (ANNEX 3) and is completely digested by pepsin leaving only the heme cofactor (ANNEX 4). Therefore, the health effects of ingesting soybean leghemoglobin should be equivalent to non-symbiotic plant hemoglobins and mammalian myoglobins, which are readily consumed in the diet.

References:

M. Punta, P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L.L. Sonnhammer, S.R. Eddy, A. Bateman, R.D. Finn Nucleic Acids Research (2012). Database Issue 40:D290--- D301

Delano WL (2007) The PyMOL Molecular Graphics System (DeLano Scientific, San Carlos, CA).

Figure 1. Structural comparison of plant hemoglobins and animal myoglobins showing that proteins adopt the same globin fold. Individual plant leghemoglobins (A-B), plant non-symbiotic hemoglobins (C-E), and animal myoglobins (F-H), are shown in ribbon representation colored in gray, heme porphyrin ring is shown in red stick representation, and iron in blue CPK representation. Superposition of individual proteins shows that the 3D structure of soybean leghemoglobin is highly similar leghemoglobins, non-symbiotic hemoglobins, and myoglobins from different species (I-L).

Figure 1. Plant hemoglobins and animal myoglobins adopt the same structural fold.

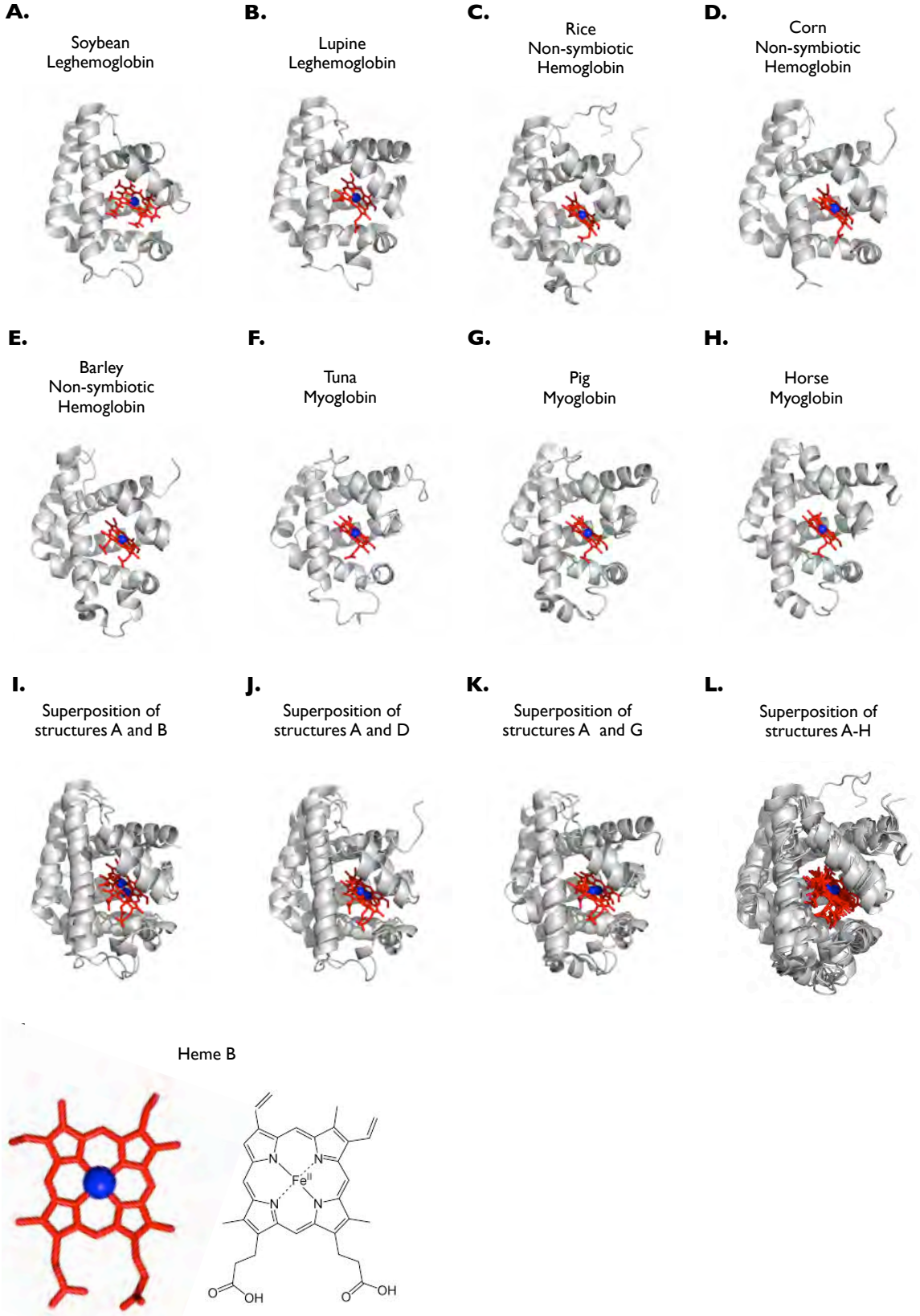


Table 1. Structural comparison between plant hemoglobins and animal myoglobins. Root-mean-square-deviation (RMSD) between all backbone atoms of superimposed X-ray crystallography protein structures (respective PDB codes are shown in parenthesis).

Species		RMSD (Å)
Soybean leghemoglobin (1BIN)	Horse myoglobin (1YMB)	4.5
Soybean leghemoglobin (1BIN)	Pig myoglobin (1PMB)	4.4
Soybean leghemoglobin (1BIN)	Tuna myoglobin (1MYT)	3.6
Soybean leghemoglobin (1BIN)	Barley non-symbiotic hemoglobin (2OIF)	2.5
Soybean leghemoglobin (1BIN)	Corn non-symbiotic hemoglobin (2R50)	1.0
Soybean leghemoglobin (1BIN)	Rice non-symbiotic hemoglobin (1D8U)	1.0
Soybean leghemoglobin (1BIN)	Lupine leghemoglobin (2GDM)	0.8
Soybean leghemoglobin (1BIN)	Soybean leghemoglobin (1FSL)	0.5

Annex 2
Dr. Richard E. Goodman Expert Opinion of Assessment of the Safety of Soy Leghemoglobin

University of Nebraska Lincoln

FOOD ALLERGY RESEARCH AND RESOURCE PROGRAM

31 October, 2013

Maraxi, Inc.
525 Chesapeake Drive
Redwood City, CA 94063

Subject: Summary of the Allergenicity and Toxicity Assessment for LegHg Protein for Food

My laboratory has performed a weight of evidence assessment of the potential allergenicity and toxicity of the heme-protein of soybean (*Glycine max*) following the principles of the Codex Alimentarius Guidelines for risk analysis of foods derived from modern biotechnology (CAC/GL 44-2003). The assessment focused on the soybean leghemoglobin protein, with the full amino acid sequence listed in the NCBI protein database as GI:126241.

Public information from peer-reviewed literature in PubMed was evaluated for evidence of allergy and toxicity associated with soybean protein known as "leghemoglobin" (and five other homologous proteins from other source organisms). Over 380 articles were found relating allergy/allergenicity or toxicity of soybean. However, refining the search to include the heme or leghemoglobin protein produced no matches. Clearly soybean is a common source of allergy and auto-annotated keywords identify many irrelevant papers. Abstracts of the 380+ papers were reviewed to ensure they were irrelevant. There is a tremendous history of safe consumption of soybeans, with the exception of food allergy, but no published evidence that the legHb is associated with allergic reactions or toxicity.

Bioinformatics searches (amino acid sequence comparisons) were performed comparing the sequence (GI:126241) with known and putative allergens in AllergenOnline.org, version 13. Using full-length FASTA alignment, which is the most predictive of cross-reactivity (modest to high risk if >50% identity over most of the full-length of the protein), there were no significant overall alignments. A search using the conservative sliding window of 80-amino acid algorithm to identify matches of >35% identity, there was only a single segment alignment (35.2% identity) in a highly gapped overlap of 105 amino acids to potato patatin (a minor allergen). The gapping and high lack of identified alignment to this protein in the full-length FASTA alignment demonstrate the match was a random match without scientific rationale to believe that a person allergic to patatin of potato would have IgE binding to, or would react with the soybean protein. In addition, there are no published reports of allergic cross-reactions with potato and soybean. In addition, a full-BLASTP search was performed against the NCBI protein database using keyword limit of "allergen" to ensure that the AllergenOnline.org database did not miss sequences that have any associated definition of "allergen". No significant alignments were found (>50% identity over most of the length of legHb). The poor scoring alignments (low percent identity, short alignment and high E scores) in the report are to proteins associated with allergenic sources, or to proteins that the NCBI auto-annotation indicated are similar to an allergenic protein.

Bioinformatics search for similarities to toxic proteins. There are no comprehensive toxin databases. Yet there are very view proteins other than venoms of snakes, insects, jellyfish and other simple oceanic organisms as well as a number of bacterial and fungal toxic that are

143 Food Industry Building / P.O. Box 830955 / Lincoln, NE 68583-0955
Richard E. Goodman: Office (402) 472-0452 Cell (402) 417-5549
FAX (402) 472-1693

known to be toxic. Their mechanisms of action between different classes of toxins vary and there are no guidelines on sequence identities that regulators generally accept as limits below which there is no concern regarding shared toxicity. However, general experiences regarding protein structure and function can be used to rule out likely shared activities for alignments that are poor relative to other known proteins. Proteins that do not share more than 30% identity and almost complete shared alignment length are unlikely to be folded the same (3-D structure) and share structure often (not always) means likely shared function. The BLASTP search of Glycine max legHb (GI:126241) with NCBI protein database using the keyword limit "toxin" did not uncover any significant alignments to any known toxic proteins. The highest scoring alignments were all less than 30% identity, with E scores higher than 1, and alignments approximately or less than half of the length of the toxic proteins. In addition, these poorly matched "toxins" do not have significant associations of toxicity except that they are produced by organisms (*Vibrio cholera*, *Escherichia coli* and *Enterococcus faecalis*) that do produce active toxins. These results do not raise concerns of potential toxicity for the soybean leghemoglobin protein.

Finally we tested the stability of the soybean hemoglobin protein in a model simulated gastric digestion study that includes fixed concentration of protein to pepsin (enzyme) activity and evaluation of digestion resistance at times up to one hour at pH 2 and 37ºC. The assay conditions that were used have been published (Ofori-Anti et al., 2008) and used to evaluate proteins in genetically modified crops and novel ingredients. There is a positive correlation between the stability of abundant dietary proteins in this assay and food allergy. In addition, proteins that are rapidly digested by pepsin are unlikely to act as toxins in the digestive tract. The soybean legHb was rapidly digested in this assay to less than 10% residual protein in less by 0.5 minutes. No residual legHb was detected after 1 minute. No stable fragments were detected either, indicating low potential risk of allergy or toxicity.

My conclusion from this "weight of evidence" approach to dietary protein safety is that the soybean leghemoglobin is very unlikely to present a risk of dietary allergy or toxicity to consumers.

(b) (6)

Richard E. Goodman, PhD, FAAAAI
Research Professor
Dept. of Food Science & Technology
University of Nebraska-Lincoln
143 Food Industry Complex
Lincoln, NE 68583-0955

Rgoodman2@unl.edu
+1 (402) 472-0452

Annex 3
Bioinformatic analysis of six heme proteins from diverse taxonomic sources for potential use in foods

Study Title

Bioinformatic analysis of six heme proteins from diverse taxonomic sources for potential use in foods

Authors

Richard E. Goodman
Kwame Andoh-Kumi

Study Completed On

30 October, 2013

(b) (6)

Performing Laboratory

Food Allergy Research and Resource Program
Food Science and Technology
University of Nebraska
143 Food Science & Technology
Lincoln, NE 68583-0955

Laboratory Project ID

Study Number: REG Maraxi 1R

1.0 Introduction

Maraxi, Inc is investigating the potential use of a heme protein as an iron-carrying protein for nutritional use in food products. The potential sources of heme proteins under investigation are from plant and bacterial sources and include: soybean (*Glycine max*), barley (*Hordeum vulgare*), mungbean (*Vigna radiata*), a bacterium from the Firmicutes lineage (*Bacillus subtilis),* a bacterium from the Chlamydiae / Verrucomicrobia group (*Methylacidiphilum infernorum*) and a bacterium from Aquificae (*Aquifex aeolicus*).

Heme proteins (sometimes referred to as hemoglobins) are ubiquitous in nature. Various forms provide different functions (oxygen carrier, catalysis of enzymatic reactions, electron transport, sensory and defense functions) related to the ecology of the source organisms. Heme proteins have been shown to be expressed in bacteria, fungi, higher plants and animals (Everse, 2004). The structure and general function of heme proteins are highly conserved throughout nature. Sprouted barley (Malted Barley) has been reported to contain 0.8% hemoglobin by mass (Duff et al., 1998). Since sprouted barley is widely used in the beverage industry (malted barley) and in the baking industry (malted barley flour), dietary exposure is common. Thus, it is clear that various heme proteins of plant and animal origin are widely consumed in the human diet.

Maraxi, Inc. has proposed using highly purified (>90% purity), food proteins processed from natural sources, or expressed as recombinant proteins by fermentation as a source of accessible iron associated with digestible proteins by incorporating the protein at from 0.1% to 1% in meat replacement products. They are interested in evaluating the potential allergenicity and toxicity of these candidate proteins to ensure product safety.

2.0 Purpose

The purpose of this study is to perform an initial evaluation or screening of the potential allergenicity and toxicity of the six purified heme proteins based on published literature about the source of the genes and bioinformatics (sequence comparisons) of proteins with known allergens and toxins. The intent is to guide decisions regarding whether additional safety tests would be needed for any of the proteins if used in food products.

3.0 Methods

3.1 Scientific literature search strategies. The PubMed database (http://www.ncbi.nlm.nih.gov/pubmed) maintained by the U.S. National Library of Medicine was used as the primary data source for scientific literature on allergy and toxicity. The primary question is whether the source of the gene is a common cause of allergy or toxicity. The data (authors, publication, date and abstracts) from searches were saved to files for review. All publication abstracts were manually reviewed and any likely relevant publications suggesting adverse health risks were investigated further by reading the journal articles.

3.1.1 Search for allergenicity. Search terms "gene source" AND "allergen" as well as "gene source" AND "allergy" were used on 28 &29 May, 2013 and repeated on 13 June, 2013.

3.1.2 Search for toxicity. Search terms "gene source" AND "toxin" as well as "gene source" AND "toxicity" were used on 28 & 29 May, 2013 and repeated on 13 June, 2013.

3.2 Sequence database search strategies.

The AllergenOnline version 13 (http://www.allergenonline.org/) and the NCBI Entrez Protein (http://www.ncbi.nlm.nih.gov/BLAST/) databases were used as the protein amino acid data sources for the sequence comparisons for allergens and toxins. The AllergenOnline database was updated in 12 February 2013 and is maintained by the Food Allergy Research and Resource Program of the University of Nebraska. Protein entries in the Entrez search and retrieval system is compiled and maintained by the NCBI of the National Institutes of Health (U.S.A.). The database is potentially updated or modified daily, and therefore the date of sequence searches by BLASTP is relevant to the dataset used in the BLASTP searches. BLASTP and FASTA3 are unique computer algorithms that provide similar local alignments and results if the appropriate scoring matrices and criteria are used.

3.2.1 FASTA3 overall search of AllergenOnline. The potential sequential and inferred structural similarities of the six heme proteins were evaluated using version 13 of AllergenOnline.org.

3.2.2 FASTA3 of AllergenOnline by 80 aa segments. This short (80-amino acid) segment search is based on the recommendation of Codex (2003). The rationale is that this might help in identifying structural motifs, much shorter than the intact protein, which might contain a conformational IgE binding epitope. It should also help to identify potentially cross-reactive proteins that are not true homologues of an allergen that have significant local identities that might provide an immunological target for IgE antibodies in those with allergies to the matched allergen. A match of >35% with a known allergen will suggest further testing for possible cross-reactivity although matches using the sliding 80 amino acid window search that are not also identified by overall FASTA search may represent an irrelevant

alignment. Thus evaluation of *E* scores and relative comparison of matched sequences with the NCBI database is sometimes warranted.

3.2.3 BLASTP of NCBI Entrez with "allergen" as keyword limit. The BLASTP is available on the NCBI Entrez website (http://www.ncbi.nlm.nih.gov/BLAST/). The current version is BLASTP 2.2.28+ (27 May, 2013). A BLASTP search was used comparing each complete query sequence against the entire Entrez Protein database, with a limit option selected to query entries for "allergen", to align only with proteins identified as allergens. The purpose of this BLASTP search is to ensure that a significant match with a newly discovered allergenic sequence that has not yet been entered into AllergenOnline is not overlooked. Evaluation of the *E* value, the length of the alignment and the percent identity of any identified match is necessary to judge the significance of any alignment using BLASTP.

3.2.4 BLASTP of NCBI Entrez without keyword limit. The purpose of this BLASTP search is to compare the six heme proteins to all known protein sequences to evaluate whether there are other similar proteins from other organisms that might provide information of safe exposure to homologues of this protein.

3.2.5 BLASTP of NCBI Entrez with "toxin" as keyword limit. The purpose of this BLASTP search is to identify matches to known toxic proteins (toxins) and if alignments share significant identities, to determine potential risks that would require further testing.

4.0 Results and Discussion. The summary results for the PubMed search using the various protein sources and search terms, and the amino acid sequences of the six heme proteins, are presented here.

4.1 PubMed Searches. The scientific literature database, PubMed, was searched for evidence that the heme proteins are likely sources of allergy or toxicity. The search did not reveal evidence that the species that the six genes were isolated from represents food safety risk.

4.1.1 Allergenicity. Summary of all information is present in Table 1.

4.1.1.1 *Aquifex aeolicus*: A search of PubMed using only the organism name, "*Aquifex aeolicus*" returned 426 articles. Restricting the search of PubMed by including both "*Aquifex aeolicus*" AND "allergen" returned no references. Further searches with terms "Aquifex aeolicus" AND "allergy" or alleg* also had no returns. The conclusion from the literature search is that allergy to *Aquifex aeolicus* has not been reported and therefore does not raise a possible allergy concern.

4.1.1.2 ***Methylacidiphilum infernorum***: A search of PubMed using only the organism name, "*Methylacidiphilum infernorum*", returned only six articles. Restricting the search of PubMed by including both "*Methylacidiphilum infernorum*" AND "allergen" returned no references. Further searches with terms "*Methylacidiphilum infernorum*" AND "allergy" or alleg* also had no references. The conclusion from the literature search is that allergy to *Methylacidiphilum infernorum* has not been reported and therefore does not raise a possible allergy concern.

4.1.1.3 ***Bacillus subtilis***: A search of PubMed using only the organism name, "*Bacillus subtilis*", returned 28,108 articles. Restricting the search of PubMed by including both "*Bacillus subtilis*" AND "allergen" returned only 29 references. None of these articles referred to an allergic response to hemoglobin or a heme protein. A second search with terms "*Bacillu subtilis*" AND "allergy" also had 141 references but further search with terms "*Bacillus subtilis*" AND "allergy" AND "hemoglobin" yielded no references. It can be concluded from the literature search that the information does not raise a concern of possible allergy to the heme proteins of *Bacillu subtilis*.

4.1.1.4 ***Vigna radiata*** **(mungbean):** A search of PubMed using only the organism name, "*Vigna radiata"*, returned 469 articles. Restricting the search of PubMed by including both "*Vigna radiata"* AND "allergen" returned only 5 references. None of these implicated an allergic response to hemoglobin or a heme protein from the mungbean. A second search with terms "*Vigna radiata*" AND "allergy" yielded 3 references that were also part of the first search results. It can be concluded from the literature search that the heme protein of *Vigna radiata* does not raise a concern of possible allergy.

4.1.1.5 Hordeum vulgare (barley): A search of PubMed using only the organism name, "*Hordeum vulgare*" returned 7,653 articles. Restricting the search of PubMed by including both "*Hordeum vulgare*" AND "allergen" returned 37 references. None of these relates an allergic reaction to hemoglobin or a heme protein in barley. A second search with terms "*Hordeum vulgare*" AND "allergy" found 44 references but further search with terms "*Hordeum vulgare*" AND "allergy" AND "hemoglobin" yielded no references. Most of the articles centered on wheat allergens and components of barley and rye that cross-react with the wheat allergen, omega-5 gliadin. A few also talked about heat shock protein (hsp 70) of barley and corn as common allergic determinants and also on bakers' asthma. It can be concluded from the literature search that the heme protein of *Hordeum vulgare* does not raise a concern of possible allergy.

4.1.1.6 ***Glycine max*** **(soybean)**: A search of PubMed using only the organism name, "*Glycine max*", returned 21,117 articles. Restricting the search of PubMed by including both "*Glycine max*" AND "allergen" returned 383 references. A second search with terms "*Glycine max*" AND "allergen" AND "leghemoglobin" also had no references. Further searches with "*Glycine max*" AND "allergy" AND "heme" yielded one reference on

impairment of carotenoid and flavonoid biosynthesis due to mutation of Arabidopsis HY1 which is not relevant to the topic under review. It can be concluded from the literature search that the leghemoglobin from *Glycine max* does not raise a concern of possible allergy.

4.1.2 Toxicity. The search of PubMed using the taxonomic organism name for all six species evaluated here (*Aquifex aeolicus, Methylacidiphilum infernorum, Bacillus subtilis, Vigna radiate, Hordeum vulgare* and *Glycine max*) along with the terms "toxin" or "toxicity" AND/or "heme" provided similar negative outcomes to that found for in the allergenicity search (summarized in Table 1). None of the publications implicated any natural protein expressed by the organisms of interest as a toxin. The majority of the papers either reported on accumulation of environmental toxins (e.g. heavy metals or pesticides) in the organisms considered as a food crop or model organism, which might expose humans to unacceptable risks or reported on adverse effects of various environmental contaminants on the health or function of cells in the organisms. Thus, based on literature search there is no reason to suspect the heme proteins produced from any of these organisms' genes would elicit a toxic effect on consumers.

Table 1 Pubmed search results (Summary table)

Protein Source	Search results			
	Source only	**Source AND heme/hemoglobin**	**Source AND allergen/allerg***	**Source AND toxin**
Aquifex aeolicus	426 articles found	10 articles found, mostly related to cytochrome activity and one related to heme synthesis enzymes	No literature found	13 articles found but none related to the protein as a toxin. Most articles were related to synthesis of the Lipid A moiety of LPS
Methylacidiphilum infernorum	6 articles found	2 articles found, none related to allergies or toxins	No literature found	No literature found
Bacillus subtilis	28108 articles found on *B. subtilis*	50 articles found, most of them addresses issues related to the biological processes or functions of truncated hemoglobin.	Had 29 articles but none suggested that heme or hemoglobin was a potential allergen. Search with *B. subtilis* and allergen AND heme/hemoglobin yielded no results	601 articles were found initially. Searching with *B. subtilis* AND toxin AND heme/hemoglobin yielded 2 articles but both were not related to toxicology of the heme protein.
Vigna radiata (mung bean)	469 articles were found on mung bean	No literature found	5 articles found but none related to hemoglobin as a potential allergen	4 articles found. None depicted heme protein as a potential toxin.
Hordeum vulgare (Barley)	7653 articles were found on Barley	37 articles	37 articles found but most were related to wheat allergens, or components of barley (gamma-3-hordein) & rye that cross-react with wheat allergen omega-5 gliadin or bakers asthma.	263 articles found, most of these articles were on mycotoxin contamination. Searching with barley AND toxin AND heme yielded no results.
Glycine max (Soy)	21775 articles were found on Soy	191 articles found on leghemoglobin. When allergen was added to the search criteria, no reference came up.	383 articles found. Further searches with heme/leghemoglobin added to the search criteria yielded no results. One article was sited when searched with Soy AND allergy AND heme but was not relevant to the topic.	385 articles found. An additional criterion of "leghemoglobin" added to the search criteria yielded no results.

4.2 Sequence comparison of the heme proteins to allergens. The amino acid sequence of each of the six heme proteins (Table 2) was compared to known allergens using both a full-length FASTA alignment search and a sliding window of 80 comparisons against AllergenOnline.org, version 13. Additionally, a BLASTP search was performed against the NCBI database using keyword search limits of "allergen" and "toxin".

Table 2 Amino acid sequence of the six heme proteins

Organism	Hemoglobin class	Native protein sequence
Aquifex aeolicus	F-hermoglobin (Fgb) GI:15605769	MLSEETIRVIKSTVPLLKEHGTEITARMYELLFSKYPKTKELFAGASEEQPKKLANAIIAYATYIDRLEE LDNAISTIARSHVRRNVKPEHYPLVKECLLQAIEEVLNPGEEVLKAWEEAYDFLAKTLITLEKKLYSQP
Methylacidiphilum infernorum	hemoglobin-like flavoprotein (Fhb) GI:501439983	MIDQKEKELIKESWKRIEPNKNEIGLLFYANLFKEEPTVSVLFQNPISSQSRKLMQVLGI LVQGIDNLEGLIPTLQDLGRRHKQYGVVDSHYPLVGDCLLKSIQEYLGQGFTEEAKAAWT KVYGIAAQVMTAE
Bacillus subtilis	truncated hemoglobin (trHb) GI:489325624	MGQSFNAPYEAIGEELLSQLVDTFYERVASHPLLKPIFPSDLTETARKQKQFLTQYLGGPPLYTEEHGHP MLRARHLPFPITNERADAWLSCMKDAMDHVGLEGEIREFLFGRLELTARHMVNQTEAEDRSS
Vigna radiata	non-symbiotic hemoglobin (nsHb) GI:377643998	MTTTLERGFTEEQEALVVKSWNVMKKNSGELGLKFFLKIFEIAPSAQKLFSFLRDSTVPLEQNPKLKPHA VSVFVMTCDSAVQLRKAGKVTVRESNLKKLGATHFRTGVANEHFEVTKFALLETIKEAVPEMWSPAMKNA WGEAYDQLVDAIKYEMKPPSS
Hordeum vulgare	non-symbiotic hemoglobin (nsHb) GI:3913789	MSAAEGAVVFSEEKEALVLKSWAIMKKDSANLGLRFFLKIFEIAPSARQMFPFLRDSDVP LETNPKLKTHAVSVFVMTCEAAAQLRKAGKITVRETTLKRLGGTHLKYGVADGHFEVTRF ALLETIKEALPADMWGPEMRNAWGEAYDQLVAAIKQEMKPAE
Glycine max	leghemoglobin (legHb) GI:126241	MGAFTEKQEALVSSSFEAFKANIPQYSVVFYTSILEKAPAAKDLFSFLSNGVDPSNPKLT GHAEKLFGLVRDSAGQLKANGTVVADAALGSIHAQKAITDPQFVVVKEALLKTIKEAVGD KWSDELSSAWEVAYDELAAAIKKAF

4.2.1 Full length FASTA3 vs. AllergenOnline. Results of the full length FASTA3 searches of the heme proteins against AllergenOnline version 13 did not identify any significant alignment with an allergen. Scoring results for the six heme proteins showing alignments with E scores less than 10 are shown in Table 3 and demonstrate no significant matches with any allergen. Their identities (%) are markedly below the level that is likely to indicate cross-reactivity (< 50% identity, Aalberse, 2000) and it is also below the 35% identity level suggested by Codex (2003) as a match that may possibly be cross-reactive. Thus, there is only a small likelihood that any of the six heme proteins are sufficiently similar to an allergen to suspect they might trigger allergic responses in allergic subjects due to cross-reactivity.

Table 3. **Overall FASTA3 search of AllergenOnline with the heme proteins.** Only the two highest scoring alignments of known and putative allergens in AllergenOnline version 13, compared to the heme proteins, using FASTA3 are listed since none of the results were significant.

Source organism	Sequence GI #	Organism	Description	Length aa	*E* score	% Identity	aa Alignment length
Aquifex aeolicus	56405054	*Chironomus thummi thummi* Flies	Globin	161	1.6	20.0	65
	121248	*Chironomus thummi thummi*	Globin	161	2.4	21.2	161
Methylacidiphilum infernorum	121256	*Chironomus thummi thummi*	Globin	151	0.17	23.5	132
	56405052	*Chironomus thummi thummi*	Globin	161	0.53	21.7	143
Bacillus subtilis	21773	*Triticum aestivum* wheat	Unnamed protein product	307	2.2	25.0	64
	170726	*Triticum aestivum* wheat	pre-alpha-/beta-gliadin A-III	282	4.0	27.8	54
Vigna radiata	56405052	*Chironomus thummi thummi*	Globin	161	0.00012	22.3	139
	56405054	*Chironomus thummi thummi*	Globin	161	0.00014	22.3	139
Hordeum vulgare		*Chironomus thummi thummi*	Globin	161	5e -007	23.3	150
	2506461	*Chironomus thummi thummi*	Globin	162	6e-007	24.8	157
Glycine max	56405052	*Chironomus thummi thummi*	Globin	161	0.0002	26.2	145
	121244	*Chironomus thummi thummi*	Globin	161	0.00035	25.9	147

4.2.1 Sliding 80-amino acid window FASTA3 vs. AllergenOnline. Results of the comparisons of the six heme protein sequences against all of the sequences in allergenonline.org version 13 database only identified a single 80 aa alignment of the Glycine max leghemoglobin to a segment of the potato allergen, patatin, GI:21512 (Table 4). That alignment does not indicate evolutionary homology or overall structural similarity as the leghemoglobin and patatin proteins did not align with an E score of less than 10 by overall FASTA3 alignment. In addition, the single "80 amino acid" alignment required insertion of 10 gaps of from one to six amino acids, thus it is really an alignment covering 105 amino acids. Few people are allergic to cooked potato and patatin has rarely been demonstrated to bind IgE from potato allergic subjects. The epitope(s) are not known for any subjects. With the poor alignment (35.2% identity, E score ~ 0.45, 10 inserted gaps of from one to six amino acids) it is unlikely that IgE from a potato allergic subject would bind to the leghemoglobin even if the aligned region of patatin contains an IgE binding epitope. Thus, there is a small likelihood that the leghemoglobin of soybean would act as a cross-reactive allergen and an even smaller chance that the other five heme proteins would exhibit cross-reactivity as none of them showed any alignment over the Codex threshold. Thus the risk of cross-reactions for allergic consumers is very low.

Table 4. Scanning 80-mer Sliding Window Search Results for leghemoglobin from Soy

Database	AllergenOnline Database v13 (February 12, 2013)
Input Query	>query MGAFTEKQEALVSSSFEAFKANIPQYSVVFYTSILEKAPAAKDLFSFLSNGVDPSNPKLT GHAEKLFGLVRDSAGQLKANGTVVADAALGSIHAQKAITDPQFVVVKEALLKTIKEAVGD KWSDELSSAWEVAYDELAAAIKKAF
Length	145
Number of 80 mers	66
Number of Sequences with hits	1

Hit #	Defline	Species	Best %ID	# Hits > 35%	Full Alignment			Links	
					E-val	%ID	length	NCBI	Details
1	gi|21512|emb|CAA27571.1| patatin [Solanum tuberosu	Solanum tuberosum	35.20%	1of66		0.00%	0	gi|21512	GO!

AllergenOnline Database v13 (February 12, 2013)

4.2.3 BLASTP of NCBI Entrez using "allergen". The full-length amino acid sequences of the six heme proteins were compared to sequences in NCBI-Entrez, which were designated as "allergen" in the NCBI database on 28 May, 2013. The aligned matches (Tables 5 to 10) were not significant as judged by the very large E score values (>0.001) and low

identity matches (25% to 30%) with partial protein alignments. The aligned proteins would not be considered homologues of the heme proteins and the probability of cross-reactivity is extremely small based on observations of Aalberse (2000) and Goodman et al. (2008).

Table 5. **BLASTP of NCBI Entrez with *Aquifex aeolicus* GI: 15605769 using the keyword "allergen".** The scoring alignments with *E* scores below 10 are shown for this heme protein vs. all proteins labeled with the keyword "allergen" in the NCBI Entrez database on 28th May, 2013, using BLASTP. The sequence identities are low and / or the length of alignments are very short, indicting unlikely homology and that the overall structure is unlikely to be similar. Thus even if the NCBI sequence is a proven allergen, there is very little likelihood of cross-reactivity by the heme protein.

Sequence GI#	Organism	Description	Length aa	E score	% identity	aa Alignment length
345499008	*Heligmosomoides polygyrus bakeri* nematode	myoglobin	176	0.035	25	101
159462966	*Chlamydomonas reinhardtii* Green algae	Predicted protein	1342	0.42	33	46
195351791	*Drosophila sechellia* flies	GM23336 protein	386	2.0	26	78
160425391	*Colias eurytheme* butterflies	Single domain major allergen 1 protein	208	2.1	48	23
333927082	*Serratia sp. AS12* enterobacteria	Alkyl peroxidase reductase (Mal allergen)	221	3.5	29	41

Table 6. **BLASTP of NCBI Entrez with *Methylacidiphilum infernorum* GI: 501439983 using the keyword "allergen".** The scoring alignments with *E* scores below 10 are shown for this heme protein vs. all proteins labeled with the keyword "allergen" in the NCBI Entrez database on 28th May, 2013, using BLASTP. Note that there were only two alignments with E scores less than 10 for this heme protein. The sequence identities are low and / or the length of alignments are very short, indicting unlikely homology and that the overall structure is unlikely to be similar. Thus even if the NCBI sequence is a proven allergen, there is very little likelihood of cross-reactivity by the heme protein.

Sequence GI#	Organism	Description	Length aa	E score	% identity	aa Alignment length
300706913	*Nosema ceranae BRL01* microsporidians	Hypothetical protein	603	1.1	34	53
29867505	*Methanohalobium evestigatum Z-7303* euryarchaeotes	Peroxiredoxin	264	1.9	25	51

Table 7. **BLASTP of NCBI Entrez with *Bacillus subtilis* GI: 489325624 using the keyword "allergen".** The scoring alignments with *E* scores below 10 are shown for this heme protein vs. all proteins labeled with the keyword "allergen" in the NCBI Entrez database on 28th May, 2013, using BLASTP. The sequence identities are low and / or the length of alignments are very short, indicting unlikely homology and that the overall structure is unlikely to be similar. Thus even if the NCBI sequence is a proven allergen, there is very little likelihood of cross-reactivity by the heme protein.

Sequence GI#	Organism	Description	Length aa	E score	% identity	aa Alignment length
340742817	*Amphioctopus fangsiao* Cephalopod	Arginine kinase	348	1.2	26	61
488389824	*Staphylococcus ingdunensis* Firmicutes	Conserved hypothetical protein	310	3.5	23	86
281203854	*Polysphondylium palladium PN500* Cellular slime mold	Hypothetical protein PPL_08696	388	4.4	29	51
2420425854	*Sorghum bicolor* Sorghum	Hypothetical protein	266	5.9	35	26
496384984	*Rhodopirellula sp. SWK7* Planctomycetes	alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	199	6.0	24	104

Table 8. **BLASTP of NCBI Entrez with *Vigna radiata* GI: 377643998 , using the keyword "allergen".** The scoring alignments with *E* scores below 10 are shown for this heme protein vs. all proteins labeled with the keyword "allergen" in the NCBI Entrez database on 28th May, 2013, using BLASTP. The sequence identities are low and / or the length of alignments are very short, indicting unlikely homology and that the overall structure is unlikely to be similar. Thus even if the NCBI sequence is a proven allergen, there is very little likelihood of cross-reactivity by the heme protein.

Sequence GI#	Organism	Description	Length aa	E score	% identity	aa Alignment length
110749448	*Apis mellifera* Honey bee	Hypothetical protein	233	4.7	39	28
340716601	*Bombus terrestris* buff-tailed bumblebee	Hypothetical protein	233	4.8	39	28
496571479	*Erysipelotrichaceae bacterium 3_1_53* firmicutes	Hypothetical protein	281	8.0	33	52
496571479	*Erysipelotrichaceae bacterium 3_1_53* firmicutes	Hypothetical protein	281	8.0	33	52
8453086	*Anisakis simplex* herring worm	paramyosin isoform	473	9.8	26	127

Table 9. **BLASTP of NCBI Entrez with *Hordeum vulgare* GI:3913789 using the keyword "allergen".** The scoring alignments with *E* scores below 10 are shown for this heme protein vs. all proteins labeled with the keyword "allergen" in the NCBI Entrez database on 28th May, 2013, using BLASTP. The sequence identities are low and / or the length of alignments are very short, indicting unlikely homology and that the overall structure is unlikely to be similar. Thus even if the NCBI sequence is a proven allergen, there is very little likelihood of cross-reactivity by the heme protein. In this case there were only three alignments with an E score less than 10.

Sequence GI#	Organism	Description	Length aa	E score	% identity	aa Alignment length
474067870	*Triticum urartu* wheat	Expansin-B12	372	1.2	43	28
152975988	*Bacillus cytotoxicus NVH 391-98* Firmicutes	alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	191	3.8	30	57
491363465	*Marichromatium purpuratum*	alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	185	6.4	37	46

Table 10. **BLASTP of NCBI Entrez with *Glycine max* GI:126241 using the keyword "allergen".** The scoring alignments with *E* scores below 10 are shown for this heme protein vs. all proteins labeled with the keyword "allergen" in the NCBI Entrez database on 28th May, 2013, using BLASTP. The sequence identities are low and / or the length of alignments are very short, indicting unlikely homology and that the overall structure is unlikely to be similar. Thus even if the NCBI sequence is a proven allergen, there is very little likelihood of cross-reactivity by the heme protein. In this case only four proteins were matched with an E score of less than 10.

Sequence GI#	Organism	Description	Length aa	E score	% identity	aa Alignment length
332669326	*Cellulomonas fimi ATCC 484* Bacteria	Pectate lyase	705	2.3	25	102
428316091	*Oscillatoria nigro-viridis PCC 7112* Cyanobacteria	SCP-like extracellular	219	7.7	35	31
297570037	*Desulfurivibrio alkaliphilus AHT2* d-proteobacteria	alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	208	7.8	31	42
257094957	*Candidatus Accumulibacter phosphatis clade IIA str. UW-1* b-proteobacteria	alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	166	7.8	48	21

4.2.4 BLASTP of NCBI without keyword limit. The full-length of the six heme proteins were compared to all sequences in NCBI-Entrez database on 28 May, 2013. The scoring alignments with *E* scores of the top 10 non-heme protein alignments identified by BLASTP on 28 May, 2013 were considered in some detail to determine if there is significant homology to proteins of sources with likely safe human exposure or unsafe (allergenic or toxic) exposure. The bacterial heme proteins were only closely related to other bacterial heme proteins, including those from bacteria that are known to caue human disease, although they are all only similar to other heme proteins, listed as likely oxygen carrying proteins. The plant heme proteins were most closely related to other eukaryotic hemes and were about 26% or more identical to some chordate heme proteins. These proteins are clearly evolutionarily related to oxygen carrying proteins from diverse sources, including organisms that humans are exposed to without harm and some organisms (certain

bacteria) that they are harmed by. Thus, the results from BLASTP comparison to all proteins were neutral, but the ubiquitous nature of heme proteins without obvious indications of harm suggest they are generally safe.

4.3 BLASTP of NCBI Entrez with "toxin". The full-length sequence of the six heme proteins were compared to sequences in NCBI-Entrez, which were designated as "toxin" in the NCBI database on 28/29 May, 2013. The top five aligned proteins in are shown for the six heme proteins. Some very high identity matches were found, with homologues of oxygen binding globins or heme proteins (e.g. 38% to 71% identity for heme proteins *Aquifex aeolicus* and *Bacillus subtilis* with small *E* score values of < 5e-19 (Tables 11 &13). These appear to be homologues of the genes from these two sources. They are identified with the BLASTP using a toxin keyword as they are from "toxic" organisms (e.g. *Vibrio cholerae*). In addition some references within the NCBI entries for the protein segments, visible in the ASN.1 view of data in the Protein view refer to toxin or anti-toxin. Therefore a search was performed with "flavohemoprotein" AND "toxin", which identified a paper that explains the connection. A key function of some of these proteins, like nitric oxide dioxygenase of Bacillus cereus and the flavohemoproteins of the same organism seems to be to detoxify nitric oxide rather than just acting as an oxygen carrier (Gardner et al., 1998). There does not appear to be a basis to suspect that the six heme proteins are likely toxins.

Table 11. **BLASTP of NCBI Entrez "toxin" with *Aquifex aeolicus* heme.** The best scoring alignments to putative toxins shown in the NCBI Entrez database on 28th May, 2013, were identified by BLASTP with the full-length sequence of the heme protein from *Aquifex aeolicus*, GI:15605769.

Sequence GI#	Organism	Description	Length aa	E score	% identity	aa Alignment length
196021452	*Bacillus cereus* Firmicute	Nitric oxide dioxygenase	402	1e-31	47	141
375283521	*Bacillus cereus* Firmicute	flavohemoprotein	402	1e-31	47	141
47554187	*Bacillus cereus G9241* Firmicute	flavohemoprotein	402	6e-31	46	141
388336311	*E. coli 0111:H11 str. cum9534* Enterobacteria	Bifunctional nitric oxide dioxygenase	163	2e-30	40	141
190901078	*E. coli B7A* Enterobacteria	flavoprotein	396	5e-29	40	141

Table 12. **BLASTP of NCBI Entrez "toxin" with *Methylacidiphilum infernorum* heme.** The best scoring alignments to putative toxins shown in the NCBI Entrez database on 28th May, 2013, were identified by BLASTP with the full-length sequence of the heme protein in *Methylacidiphilum infernorum* GI:501439983.

Sequence GI#	Organism	Description	Length aa	E score	% identity	aa Alignment length
196021452	*Bacillus cereus 03BB10B* Firmicute	Nitric oxide dioxygenase	402	5e-16	31	135
375283521	*Bacillus cereus NC7401* Firmicute	flavohemoprotein	402	5e-16	31	135
47554187	*Bacillus cereus G9241* Firmicute	flavohemoprotein	402	1e-16	31	135
47554187	*Bacillus cereus G9241* Firmicute	flavohemoprotein	402	1e-16	31	135
262028231	*Vibrio cholerae INDRE 91/1* g-proteobacteria	flavohemoprotein	394	2e-12	26	133

Table 13 BLASTP of NCBI Entrez "toxin" with *Bacillus subtilis*. The best scoring alignments to putative toxins shown in the NCBI Entrez database on 28th May, 2013, were identified by BLASTP with the full-length sequence of the heme protein in *Bacillus subtilis*, GI:489325624.

Sequence GI#	Organism	Description	Length aa	E score	% identity	aa Alignment length
375283270	*Bacillus cereus NC7401* Firmicute	Globin protein	132	1e-58	71	124
196024904	*Bacillus cereus 03BB108* Firmicute	Protozoan globin protein	132	3e-58	71	124
47554951	*Bacillus cereus G9241* Firmicute	Globin protein	132	3e-58	71	124
397654575	*Corynebacterium ulcerans 0102*	Hemoglobin-like protein	130	5e-19	38	128
497730520	*Gemmate obscuriglobus* Aquatic bacteria	Hypothetical protein	139	1.5	29	63

Table 14. **BLASTP of NCBI Entrez "toxin" with *Vigna radiata* heme.** The best scoring alignments to putative toxins shown in the NCBI Entrez database on 28th May, 2013, were identified by BLASTP with the full-length sequence of the heme protein in *Vigna radiata*, GI:377643998.

Sequence GI#	Organism	Description	Length aa	E score	% identity	aa Alignment length
375283521	*Bacillus cereus NC7401* Firmicutes	flavohemoprotein	402	0.001	37	57
47554187	*Bacillus cereus G9241* Firmicutes	flavohemoprotein	402	0.001	37	57
196021452	*Bacillus cereus 03BB108* Firmicutes	nitric oxide dioxygenase	402	0.001	37	57
262028231	*Vibrio cholerae INDRE 91/1* g-proteobacteria	flavohemoprotein	394	0.003	19	137
262023727	*Vibrio cholerae RC27* g-proteobacteria	flavohemoprotein	394	0.003	19	137

Table 15. **BLASTP of NCBI Entrez "toxin" with *Hordeum vulgare* heme.** The best scoring alignments to putative toxins shown in the NCBI Entrez database on 28th May, 2013, were identified by BLASTP with the full-length sequence of the heme protein in *Hordeum vulgare*, GI:3913789.

Sequence GI#	Organism	Description	Length aa	E score	% identity	aa Alignment length
262023727	*Vibrio cholerae RC27* g-proteobacteria	flavohemoprotein	394	0.005	19	139
262028231	*Vibrio cholerae INDRE 91/1* g-proteobacteria	flavohemoprotein	394	0.005	19	139
47554187	*Bacillus cereus G9241* Firmicutes	flavohemoprotein	402	0.033	38	50
375283521	*Bacillus cereus NC7401* Firmicutes	flavohemoprotein	402	0.035	36	56
196021452	*Bacillus cereus 03BB108* Firmicutes	nitric oxide dioxygenase	402	0.035	38	50

Table 16. **BLASTP of NCBI Entrez "toxin" with *Glycine max* heme.** The best scoring alignments to putative toxins shown in the NCBI Entrez database on 28th May, 2013, were identified by BLASTP with the full-length sequence of the heme protein in *Glycine max* GI:126241.

Sequence GI#	Organism	Description	Length aa	E score	% identity	aa Alignment length
262028231	*Vibrio cholerae INDRE 91/1* g-proteobacteria	Flavohemoprotein	394	2.9	20	136
388345913	*Escherichia coli O103:H25 str. CVM9340* Enterobacteria	hypothetical protein ECO9340_25708	192	3.2	27	113
262023727	*Vibrio cholerae RC27* g-proteobacteria	Flavohemoprotein	394	3.3	20	136
300849370	*Enterococcus faecalis TUSoD Ef11* Firmicutes	putative alkyl hydroperoxide reductase F subunit	560	5.8	28	53
188487268	*Escherichia coli 53638* Enterobacteria	Conserved hypothetical protein	192	7.3	27	113

4.3 Bioinformatics summary for the heme proteins. None of the results from the bioinformatics searches of the heme protein amino acid sequences, compared to known and putative allergens or toxins, suggested any clear and significant sequence similarity that suggests potential adverse effects for consumers. The leghemoglobin of *Glycine max* scored closest to significance in the comparison to allergens by sliding 80 amino acid FASTA match with AllergenOnline.org. There was a single alignment of 80 amino acids with 35.2 %, just over "threshold" of concern according to Codex (2003). The query protein length is 145 amino acids. Patatin has a full-length sequence is 386 amino acids. No alignment is found between these two proteins when full length FASTA is run with AllergenOnline.org or BLAST with NCBI Protein database. The single alignment using the sliding window of 80 requires insertion of 10 gaps in the sequences to allow alignment and the E score is between 0.4 and 0.5 (varies each time due to the control shuffle calculation in the FASTA program). When the *E* score is greater than ~ 1e-7, there is rarely a match of significance in predicting allergenic cross-reactivity However, the alignment to patatin from *Solanum tuberosum* (potato), a rarely allergenic food with few reports of IgE binding to the protein, . There was a single match using the sliding window of 80, but that match required a total of 10 gaps of from 1 to six amino acids and the alignment is for less than 1/3 of the length of patatin. The protein is therefore not expected to have shared three-dimensional structure or sufficient shared linear sequence to assume IgE cross-reactivity. The likelihood of cross-reactivity for that protein is still very low and the protein is not known to be allergenic. The search results from BLASTP on toxicology suggest that none of the six heme proteins are likely toxins. Regarding potential toxicity, the heme proteins of *Aquifex aeolicus* and *Bacillus subtilis* do show high sequence identities and low *E* score values to homologous heme proteins from "toxic" organisms, and the proteins were identified using "toxin" as a keyword. Yet no evidence was found to suggest that these proteins would have a toxic effect on mammalian consumers. In addition, there were no significant similarities between the gene sequences from the other four sources (*Methylacidiphilum infernorum*, *Vigna radiata*, *Hordeum vulgare and Glycine max)* and known toxins in the database.

5.0 Conclusions

Bioinformatics analyses were performed on six specific heme proteins from soy (*Glycine max*), barley (*Hordeum vulgare*), mungbean (*Vigna radiata*), *Bacillus. subtilis, Methylacidiphilum infernorum* and *Aquifex aeolicus* to evaluate whether there might be some safety concerns for foods produced with these proteins included as ingredients.

Based on the evidence and my knowledge of cross-reactive IgE binding, there does not seem to be a need to perform serum IgE binding studies to compare leghemoglobin of *Glycine max* to patatin of potato as might be expected if the proteins were clear homologues or shared much more significant identities (Goodman, 2008).

Sequence comparisons of the six heme proteins to known toxins identified some significant identity matches to homologous heme proteins from bacterial species that are pathogenic or toxic, yet the homologous proteins have not been identified as toxins. Thus, *in vitro* and *in vivo* toxicity testing should not be required to evaluate food safety in our opinion.

6.0 References

Aalberse RC. (2000). Structural biology of allergens. *J Allergy Clin Immunol* 106:228-238.

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.

Duff SMG, Guy PA, Nie X, Durnin DC, Hill RD. (1998). Haemoglobin expression in germinating barley. *Seed Science Research* 8(4):431-436.

Everse, J. (2004). Heme Proteins. In *Encyclopedia of Biological Chemistry, Volume 2* (pp. 354-361). Elsevier Inc. *Ed*. WJ Lennarz and MD Lane. First edition, 2004.

Gardner PR, Gardner AM, Martin LA, Salzman AL. 1998. Nitric oxide dioxygenase: An enzymic function for flavohemoglobin. Proc Natl Acad Sci USA 95:10378-10383.

Goodman RE. (2008). Performing IgE serum testing due to bioinformatics matches in the allergenicity assessment of GM crops. *Food Chem Toxicol* 46:S24-S34

Goodman RE, Vieths S, Sampson HA, Hill D, Ebisawa M, Taylor SL, van Ree R. (2008). Allergenicity assessment of genetically modified crops—what makes sense? *Nat Biotechnol* 26(1):73-81.

Annex 4
Leghemoglobin *in vitro* digestibility study in human simulated gastric fluid

FINAL REPORT

STUDY TITLE

LegHb: *in vitro* digestibility study in human simulated gastric fluid (pH 2.0)

AUTHORS

Richard E. Goodman
(b) (6)
(b) (6)

20 Oct 2013

STUDY COMPLETED ON

20 October 2013

PERFORMING LABORATORY

Goodman Laboratory
University of Nebraska
Dept. of Food Science & Technology
143 Food Industry Complex
Lincoln, NE 68583-0955, USA

SUBMITTERS/SPONSORS

Maraxi, Inc.
525 Chesapeake Drive
Redwood City, CA USA

LABORATORY STUDY ID

REG-2013-Pepsin LegHb

SUMMARY

The Leghemoglobin (LegHb) protein used in this study was produced in and purified from *E. coli* and supplied by Maraxi, Inc. of Redwood City, CA. The gene was synthesized to encode Lgb2 from the soybean *Glycine max* (145 amino acids, UniProt P02236) and codon-optimized for expression in *E. coli*. The test material was supplied as a concentrated aqueous solution from the study sponsor (Maraxi, Inc.). Maraxi Inc. ensures the LegHb protein identity and purity being >90%.

The aqueous LegHb solution was then subjected to digestion in pepsin based on the protocol in Thomas *et al*. (2004), as refined by Ofori-Anti et al., 2007. The time to reach 90% digestion of the protein by pepsin was estimated as the first sample time having less than 10% residual protein compared to diluted non-digested sample protein. The ability of the assay to detect 10% residual protein was determined prior to the digestion tests using serial dilutions of the test protein in a SDS-PAGE, Coomassie blue staining to ensure that a residual of 10% undigested control sample detectable under the conditions used for the study. The primary LegHb band migrated at ~ 16 kDa in SDS-PAGE and a slightly lower MW secondary band was clearly visible at ~13 kDa that is likely to represent a natural proteolytic product of LegHb from the *E. coli* culture. Pepsin was diluted in simulated gastric fluid (SGF) with the pH adjusted to 2.0. On each assay day the pepsin solution was tested for proteolytic activity by digestion of hemoglobin. The mass ratio of pepsin to LegHb was adjusted to achieve ~ 10 units of pepsin activity per microgram of test protein. Two digestion assays in SGF buffer at pH 2.0 were performed on different weeks. Digestions were performed at 37°C under timed conditions. Samples of the digestion mixtures were removed and neutralized at various time points (from 30 seconds to 1 hour) and samples of each were electrophoresed in SDS-PAGE and stained with Commassie blue to evaluate digestion completeness.

The results of this study demonstrated that the *E. coli*-produced LegHb protein was rapidly digested in pepsin at pH 2.0. The SDS-PAGE Coomassie blue gel staining method demonstrated that more than 90% of the *E. coli*-produced LegHb protein was digested in less than 2 minutes in replicate assays. No degradation bands were found to result from digestion of the LegHb protein. However, a faintly stained diffuse band was detected in both replicate assays at approximately 22 kDa. Proteomic analysis by trypsin digestion and LC-MSMS identified peptides corresponding to porcine pepsin as well as relatively low confidence matches to two 22 kDa undefined soybean proteins and peptides from three *E. coli* proteins. Those identities do not alter the conclusion that the *E. coli* produced LegHb is rapidly digested in pepsin at pH 2 and that no pepsin-stable fragments of LegHb were identified in the assay. Based on Codex (2003) guidelines for the allergenicity assessment, there is no added concern of risk from the presence of LegHb in food.

Table of Contents

Section **Page**

Title Page 1

Summary 2

Table of Contents 3

Abbreviations 5

1. Introduction 6

2. Materials 7
   - 2.1 Test Substance 7
   - 2.2 Control Substance 7
   - 2.3 Reference Substance 8
   - 2.4 Characterization of Test, Control, and Reference Substances 8
   - 2.5 Critical Analytical Reagents ..... 8 !

3. Test System 8
   - 3.1 Justification for Selection of the Test System 10
   - 3.2 Experimental Controls 10
   - 3.3 Sample Retention 10

4. Detailed Study Design 11
   - 4.1 Verification of Detection System Specificity and Sensitivity 11
   - 4.2 Preparation of SGF Plus pepsin 11
   - 4.3 Pepsin Activity Assay 11
   - 4.4 Control Substance Digestion 12
   - 4.5 Test Protein Digestion 12
   - 4.6 SDS-PAGE Gel 13
   - 4.7 Image Analysis 13
   - 4.8 Proteomic LC-MSMS Identity of a Faint 22 kDa Band ..... 13 !

5. Results & Discussion 14
   - 5.1 Limit of Detection 14
   - 5.2 Pepsin Activity 14
   - 5.3 Control Substance Digestion Results 14
   - 5.4 Test Protein Digestion Results 14

6. Conclusions.....15

7. References.....15

Figures & Tables

Figure 1. EZBlue Stained SDS-PAGE Gel Showing the Limit of Detection of Purified *E. coli* Produced LegHb. 17

Figure 2. ! Representative EZBlue Stained SDS-PAGE Gel Showing the Digestion of BSA Control in Simulated Gastric Fluid. (pH 2.0) 18

Figure 3. ! Representative EZBlue Stained SDS-PAGE Gel Showing the Digestion of ! Ovalbumin Control in Simulated Gastric Fluid. (pH 2.0) 19

Figure 4 Representative EZBlue Stained SDS-PAGE Gel Showing the Digestion of LegHb Test Substance in Simulated Gastric Fluid. (pH 2.0).....20

Abbreviations

aa !	amino acid
$A_{280\ nm}$!	Absorbance of light at a wavelength of 280 nm
BSA	Bovine serum albumin
CFR	Code of Federal Regulations
D0-60 !	Digestion samples (hemoglobin plus pepsin) from time 0 (quenched prior to digestion) to time 60 min
ECL	Enhanced chemiluminescence
E0	Experimental control pepsin without the hemoglobin, time 0
E60 !	Experimental control pepsin without the hemoglobin, 60 min
Hb	Acidified 2% hemoglobin
HRP !	Horseradish peroxidase
kDa	kilodalton
LegHb	Hemoglobin from *Glycine max* (Soy), *E. coli* produced
LOD	Lower limit of detection
ma	milliampere
mg !	milligram
ml !	milliliter
mM !	millimolar
µl	microliter
na	Not applicable
ng !	nanogram
NFDM	Nonfat dried milk
Ova	Ovalbumin
P0	Experimental control hemoglobin without pepsin, time 0
P60 !	Experimental control hemoglobin without pepsin, 60 min
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PFTE	Polytetrafluoroethylene
P/N	Product number, same as catalog number
P1/10 !	Experimental Control protein HPPD 10% sample for gels
R^{2} !	Square of the correlation coefficient
SDS	Sodium dodecyl sulfate
SGF	Simulated gastric fluid (without pepsin)
SOP	Standard operating procedure
T !	Time point
TCA	Trichloroacetic acid
Tris	Tris(hydroxymethyl)aminomethane
vdc	volts, direct current
v/v !	solute volume to solution volume
w/v	solute weight to solution volume

1. Introduction

Maraxi, Inc. of Redwood City, CA is developing a potential food product that will contain a purified hemoglobin protein from soybean (*Glycine max*), called leghemoglobin (LegHb). Maraxi, Inc. requested that we perform tests and an evaluation of the potential allergenicity of the LegHb protein in order to consider whether there is a risk of food allergy associated with consumption of the protein. This report describes the rationale and methods for testing the protein in an in vitro digestion assay and presents data and interpretation of the data relative to potential risks of food safety.

The *Codex Alimentarius* Commission guidelines for assessing the allergenicity of GM plants (2003) recommends assessing the introduced protein for stability in pepsin at acidic pH as an assay to help evaluate whether the introduced protein is likely to either increase the rate of sensitization to the host crop, or increase the likelihood of eliciting an allergic response in food allergic consumers. The pepsin stability assay is one study in a weight of evidence approach intended to assess the potential allergenicity of genetically modified crops (Codex, 2003). The test method for the assessment was first described by Astwood *et al.* (1996). The assay is not meant to predict whether a given protein will always be digested in the stomach of the human consumer, but does provide a simple *in vitro* correlative assay to evaluate protein digestibility. Investigation of proteins that have been tested suggest a marked positive predictive value that food allergens causing systemic reactions are relatively stable in the assay, while non-allergenic food proteins are typically digested relatively quickly (Bannon *et al.*, 2002). Purified porcine pepsin has been used to evaluate the stability of a number of food allergens and non-allergenic proteins in a multi-laboratory study that demonstrated the rigor and reproducibility in nine laboratories (Thomas *et al*., 2004). Porcine pepsin is an aspartic endopeptidase with broad substrate specificity. Pepsin is optimally active between pH 1.2 and 2.0, but inactive at pH 3.5 and irreversibly denatured at pH 7.0 (Collins and Fine, 1981; Crevieu-Gabriel *et al*., 1999). The assay is performed under standard conditions of 10 units of pepsin activity per microgram of test protein. A relatively pure form of pepsin was used Worthington Biochemical Co., pepsin A, product LS003319.

The original assay described by Astwood *et al.* (1996) recommended performing the digestion at pH 1.2, however, the FAO/WHO (2001) suggested using two pH conditions (pH 1.2 and pH 2.0). In comparing pH 2.0 vs. pH 1.2, Thomas *et al*. (2004) showed that protein digestion at pH 2.0 resulted in slightly slower rates of full-length protein and fragment degradation, but did not alter the overall sensitivity of a protein to digestion. Results at pH 1.2 were more consistent than at pH 2.0, with 91% and 77% agreement between laboratories, respectively. However, more recently, we have digested a number of proteins at both pH 1.2 and 2.0 and have not seen any significant differences (Ofori –Anti et al., 2008). Therefore in this study we only evaluated stability at pH 2.0.

The assay is performed at 37°C and samples are removed at specific times and the activity of pepsin is quenched by neutralization with carbonate buffer and Laemmli loading buffer, then heating to more than 70°C for 3 to 5 minutes. The timed digestion samples are separated by

SDS-PAGE and stained with Coomassie or colloidal blue to evaluate the extent of digestion. A review of the digestibility assay by Bannon *et al.* (2002) and by Thomas *et al.* (2004) indicates that most of the non-allergenic food proteins that have been tested are digested in around 15 to 30 seconds, while major food allergens are stable, or produce pepsin-stable fragments that are visible for from eight to 60 minutes.

Assay parameters that have not been documented in some publications include verification of pepsin activity, the limit of detection of the protein in the stained gel and an objective measurement of the time of digestion (Ofori-Anti et al., 2008). In this study, we have tested the activity of the pepsin in SGF on each day of assay using a test to evaluate digestion of bovine hemoglobin, as described by Worthington, to ensure that it is within a tolerance interval reported by Worthington for that lot of enzyme. The results of our activity assay do not exactly duplicate the labeled activity determined by Worthington, even though we use similar procedures, and we have an acceptance criterion of the Worthington certified activity, plus or minus 1,000 activity units per mg of pepsin. A second important criterion that we have included in our standard operating procedure (SOP) is an objective measured level of residual test protein (LegHb in this case) that must be reached in determining the time of digestion. We have defined the extent of digestion as 90% and determine the sample time-point when the residual is less than or equal to 10% of the amount of test protein in the initial sample. To accomplish that a dilution series of test protein is tested in the same SDS-PAGE and colloidal blue staining system as the digests are analyzed with to evaluate a limit of detection (LOD). The LOD must be lower than 10% to perform this assay. The analytical gel for the pepsin digests includes a 10% test protein sample mixed with quenched pepsin (high pH, to avoid digestion). Details and results of the study are reported here.

2. Material

2.1 Test Substance

The test substance for this study was Leghemoglobin from soybean (*Glycine max)*. The hemoglobin sample was provided by Maraxi from a fermentation of *E. coli* containing an expression plasmid with a cloned LegHb gene. The protein sample was in solution in a 15 ml screw cap disposable polypropylene centrifuge tube, shipped on ice packs. The LegH was labeled as ~ 4 mg/ml. The buffer indicated by Maraxi was 20 mM K_2HPO_4/KH_2PO_4, pH 7.4, 100 mM NaCl as labeled by Maraxi. The concentration was evaluated in our lab using Bradford assay and determined to be 4.7 mg/ml. The solution was stored at 4 °C.

2.2 Control Substance

The control substances for this study were BSA and ovalbumin. Each was tested in separate digestion assays to demonstrate the validity based on previous tests and results. The control substance tests were performed prior to the testing the samples.

2.3 Reference Substance

There was no reference substance for this study. Analytical reference standards (e.g., molecular weight markers) used in this study were documented in the data and are described in this report.

2.4 Characterization of Test, Control, and Reference Substances

Characterization of the *E. coli*-produced LegHb protein was the responsibility of Maraxi, Inc. Maraxi shared the molecular weight and the protein (amino acid sequence) with us prior to the study, which was important in analyzing results.

2.5 Critical Analytical Reagents

- Pepsin A, Worthington Biochemical Corporation, product #3319, lot #M8C10390, certified as having 2,730 activity units per mg solid
- SGF with pepsin: A 35 mM NaCl solution is adjusted in pH to 2.0 as measured with a calibrated pH meter, using 6 N HCl, to a final normality of approximately 0.084
- SGF plus pepsin: Dissolved the mass of powdered pepsin in SGF to achieve a final activity of 4,000 units per 1.52 mL of SGF, based on the activity units from Worthington
- Bovine Serum Albumin (BSA) from Sigma Chemical Co., product #A9647-100G, lot #051M1873V
- Ovalbumin, from Worthington Biochemical Corporation, product #3054, lot #52P13864.
- Hemoglobin from bovine blood, Sigma Chemical Co., product #H-2625 -25 G, lot #SLBD8821V is used to test protein pepsin activity
- EZBlue™ Gel Staining Reagent, Sigma Chemical Co., product #G-1041 -500ML, lot #SLBD4038V
- Pepsin quenching solution: 200 mM $NaHCO_3$, pH 11
- 6X Laemmli buffer, Boston BioProducts, CAS# BP-111NB
- B-mercaptoethanol, BioRad #161-0710, lot #210007236
- 5X reducing Laemmli buffer (83.3% 6X Laemmili plus 16.7% β-mercaptoethanol,), gel loading-buffer
- Precision Plus Protein™ Dual Xtra Standards from BIO-RAD (product #161-0377)
- Novex 10-20% tris-glycine polyacrylamide gels, 1.5mm thick, 15 wells (Invitrogen EC61385BOX)
- Electrophoresis running buffer
- EZBlue gel staining reagent (Coomassie stain), Sigma Chemical, product #G1041, lot #101M4337

3. Test System

The test system for this study was an *in vitro* digestion model using pepsin in simulated gastric fluid (SGF). Standard Operating Procedures (SOPs) for preparation of the SGF, determination of the detection limit assay, pepsin activity assay, digestion assay, SDS-PAGE and gel staining are on record in the laboratory. The SGF preparation and digestion procedures

were based on the methods described by Thomas *et al*. (2004) as modified by Ofori-Anti et al., (2008).

The pepsin activity assay was based on the method described by Worthington for determining the activity of pepsin. An appropriate mass of pepsin powder was dissolved in prepared SGF, pH 2.0 to provide 0.3 mg/ml as a 10 x stock, which was then diluted to 1 x with SGF. Acidified bovine hemoglobin (2% mass to volume) was prepared and digestions to evaluate the labeled pepsin activity were performed in triplicate (1.5 ml per tube). After 15 min digestion at 37ºC, samples were mixed with 5% trichloroacetic acid (TCA) for 5 mins, by adding 2.5 ml of 5% TCA to each tube, tubes were inverted several times, mixed at 37ºC, then the precipitate (undigested protein) was separated by filtration with 0.45 micron PFTE filters from the supernatant (soluble amino acids). The absorbance of the cleared supernatant was measured at 280 nm, value of the sample blanks (undigested acidified hemoglobin) absorbance was subtracted and the activity was calculated as per SOP. If the activity of the pepsin in SGF was within the specifications of our SOP (+/- 23% of the activity stated by Worthington), the labeled activity of Worthington assumed to be correct and that value was used to prepare the proper ration of pepsin to test protein (below).

The amount of pepsin powder used to prepare SGF was calculated from the specific activity on the product label was 2,730 units /mg solid pepsin product. One unit activity is defined as a change in $A_{280\ nm}$ of 0.001 at 37 °C, measured as trichloracetic acid (TCA)-soluble products using hemoglobin as the substrate. The assay was designed for fixed volumes and a fixed amount of test protein so the amount of pepsin diluted in SGF is adjusted to provide the appropriate ratio of 10 units of pepsin activity per microgram of test protein in the digestion mixture. The appropriate amount of solid pepsin was added to SGF to provide 4,000 units per 1.52 mL of SGF. The pepsin/SGF reaction mixture was preheated to 37°C in a water bath before adding 80 microliters of test protein (5 mg/mL) for a total volume of 1.6 mL, providing 10 units per microgram test protein.

Once the pre-heated (37ºC) test protein solution was mixed with pre-heated pepsin-SGF, equal volume samples were withdrawn at predetermined times (between 0.5 and 60 minutes) and added to sample tubes containing neutralization (carbonate buffer, pH 11) and denaturing reagents (reducing Laemmli buffer), which stopped the digestion. Samples were then heated to > 85°C before running in SDS-PAGE. All samples from a single digestion were applied to wells of the same SDS-PAGE gel along with molecular weight markers, undigested test protein equivalent to the initial undigested test protein sample and a 10% test protein sample and pepsin alone (to assess pepsin stainable protein bands). Samples were separated by electrophoresis, fixed for one hour (10% acetic acid, 50% methanol in water) , stained with EZBlue (at least 2 hours), destained in water and the stained image captured using a Gel Logic 440 system (Carestream, Rochester, NY). The stability of the protein was defined as the time required to achieve 90% digestion, which was estimated based on the shortest time-digested sample with a band intensity equal to, or less than the 10% undigested standard well (C0). Any new bands above approximately 3,000 MW, which were generated as intermediate products of digestion, were noted as stable (or partially stable) intermediate proteolytic fragments and were considered

based on stability. If those bands were also in the pepsin only controls (time 0 and time 60 mins), they were discounted as being from pepsin. Otherwise they would be analyzed by proteomic methods to determine whether they were fragments of the test protein.

Proteins with more than 10% stainable full-length protein band remaining at 60 minutes were considered stable. Proteins reduced to < 10% stainable band at 5 to 30 minutes were considered of intermediate stability. Proteins reduced to < 10% stainable band by 2 minutes were considered labile (rapidly digested).

3.1 Justification for Selection of the Test System

In vitro digestion models are used commonly to assess the digestibility of ingested substances. Previous studies have used this simple, *in vitro* assay to evaluate potential risk of food allergy, and demonstrated that digestibility is an important risk factor for food allergy, which might be related to initial sensitization or to elicitation once the individual is sensitized (Astwood *et al.*, 1996 and del Val *et al.*, 1999). The FAO/WHO (2001) suggested conducting the pepsin digestion assay at pH 1.2 and pH 2.0. We have performed additional independent tests showing similar results (Ofori-Ant *et al.*, 2008). In this analysis, digestion was performed at pH 2.0 as a conservative approach as some authors have claimed a lack of predictive value for the digestion assay in pepsin (Fu *et al*., 2002; Yagami *et al*., 2000). However, Bannon *et al.* (2002) reviewed a broad range of published representative pepsin digestion studies and found a strong positive predictive value when comparing the stability of allergenic and non-allergenic dietary proteins. As defined by Codex (2003), this assay is not meant to be a stand-alone determinant in evaluating the potential allergenicity of proteins introduced into GM crops, but the results are to be judged in a weight of evidence approach by regulators.

3.2 Experimental Controls

Controls in this study were meant to ensure assay reliability and include:

- Measurement of the activity of pepsin in SGF.
- Evaluation of the sensitivity of the staining properties of the test protein from serially diluted samples, in a separate, but similar SDS-PAGE gel.
- Inclusion of samples of pepsin without test protein at times zero and 60 minutes to determine whether any stainable protein bands observed in digestion samples with test protein are from the test protein, contaminants in pepsin or from pepsin autocatalysis.
- Inclusion of protein in SGF without pepsin at times zero and 60 minutes to evaluate the effect of acid and heat alone.

3.3 Sample Retention

Samples of test protein and digested samples were numbered to distinguish assay time points and assay replicates by date. Residual samples were stored at 4 °C and will be discarded approximately six months after the completion of the study.

4. Detailed Study Methods

This study evaluated the stability of leghemoglobin from Glycine *max* using pepsin at pH 2.0. A number of control steps were performed to ensure study validity. A detailed description of the study is presented here. Laboratory records and protocols are on file in the Goodman laboratory, Dept. of Food Science & Technology, University of Nebraska, Lincoln, USA.

4.1 Verification of Detection System Specificity and Sensitivity. A dilution series of LegHb was prepared with sample quantities loaded in two identical gels using 1 x reducing Laemmli buffer, covering the range representing 100% test protein per well (139 µg/ml of LegHb) down to 1.25 % (1.7 µg of LegHb). Bio-Rad precision plus protein MW markers were applied to separate lanes. Following electrophoresis, the gel was fixed and stained with EZBlue for at least 6 hours. The gel was destained 3 times with an excess of water, 15 min each time, or until the background was clear. The image captured using a Gel Logic 440 Image Station.

4.2 Preparation of SGF Plus Pepsin. The simulated gastric fluid (SGF) reaction buffer was prepared by adding 122.8 mg of NaCl to 59.2 mL of distilled water. The pH of the solution was adjusted to pH 2.0 using 6 N HCl and water. The HCl content was approximately 0.084 N, and the salt concentration was 35 mM NaCl. The certified activity of pepsin A from Worthington was used to calculate the amount of solid pepsin that was dissolved in 1.52 mL of SGF. The target was 4,000 units of activity. For this lot, the certified value was 2730 units per mg of pepsin solid material. Based on the Worthington analysis, the concentration of pepsin A used in the assay was 0.96 mg per ml, which is 0.0096 g of solid pepsin adding to 10 ml of SGF. After mixing to dissolve the pepsin it was stored at 4°C and assayed for activity and used within 24 hours.

4.3 Pepsin Activity Assay. Each time SGF plus pepsin was prepared for a digestion assay; the activity of the pepsin and the digestion assay were both completed within 24 hours. The purpose of performing the activity assay was to ensure that the pepsin was active within a pre-defined range around the certified claim of activity by Worthington. This product typically has an activity of approximately 2,740 units per mg of solid material. The activity assay we used was similar, but not identical to that used by Worthington. The tolerance was +/- 23% of the target units per mg compared to the Worthington certified claim. The SGF plus pepsin was freshly prepared and stored at 4°C just before use, and then warmed to 37°C before the addition of the target protein. The procedure was performed as follows:

4.3.1 A solution of 2% acidified bovine hemoglobin (Hb) was prepared by dissolving 0.5 g of hemoglobin (Sigma # H2625) in 20 mL of distilled water, then mixing with 5 mL of 300 mM HCl.

4.3.2 Three polypropylene screw-top centrifuge tubes were labeled as Test (#1-3), three were labeled as Blank (#1-3), each received 1.25 mL of 2% acidified Hb and all were preheated to 37°C for 10 min.

4.3.3 At a timed interval (~ 1 min.), each of the test tubes in turn received 0.25 mL of SGF plus pepsin, was mixed by gentle vortex and returned to the incubator. As each test tube reached 10 min. incubation time, 2.5 mL of 5% TCA (Sigma 6.1 N product T0699, diluted 1:20 with distilled water) was added to stop the reaction, the tube was mixed briefly by multiple inversion and then placed on ice to cool down. Then insoluble material (undigested hemoglobin) was removed using syringes (LuerLok BD 309604, 10 ml) and syringe filters (Fisher Biotech product 09-719H, 25 mm 0.45 μm PTFE).

4.3.4 Blank tubes were interspersed with the Test tubes. Blank tubes (with 1.25 mL of Hb) received 2.5 mL of 5% TCA, multiple inversion, then 0.25 mL of SGF plus pepsin. After 10 min incubation time, these tubes were also placed on ice and then filtered to remove insoluble material.

4.3.5 The absorbance at 280 nm was measured on a spectrophotometer. The activity units of pepsin per mL were calculated as the mean net absorbance (A280 nm Hb – A280 controls) multiplied by a conversion factor of 1,000 to yield units of activity per mg of solid pepsin.

4.4 Control Protein Digestions (BSA and Ova). Bovine serum albumin (BSA) and ovalbumin (Ova) digestion assays were tested as control proteins to verify the appropriate activity of the test system.

4.5 Test Protein Digestion. The soybean leghemoglobin sample (LegHb) concentration was estimated as 4 mg/ml LegHb, at 85% purity by Maraxi, and was determined to be 4.7 mg/ml total protein by Bradford assay in our laboratory. We have used that 4.7 mg/ml value for calculating concentrations. Protein solutions were kept at 4°C right before use.

4.5.1 ! Sample tube preparation. 1.5 mL centrifuge tubes were labeled as P1/10, P0, P60, D0, D0.5, D2, D5, D10, D20, D30, D60, E0, E60.

4.5.2 70 μL of pepsin quenching solution (carbonate buffer) and 70 μL of 5X Laemmli, reducing buffer were added to each tube in 4.4.1.

4.5.3 ! An aliquot of leghemoglobin in a tube labeled as P, was prepared.

4.5.3 $P_{1/10}$: 190 μL of SGF plus pepsin was added, quick heated at 85°C, then 10 μL 1/10 diluted leghemoglobin solution was added. Solution was vortexed and then heated at 85°C for 10 min.

4.5.4 Label a tube P_{mx} (no pepsin, protein control): 80 μL out of tube P and then1.52 mL SGF were added and mixed.

4.5.4.1 Immediately 200 μL into the P0 tube were removed, mixed and heated at 85°C for 10 min.

4.5.4.2 After 60 minutes at 37°C water bath, 200 μL into the P60 tube were removed, mixed and heated at 85°C for 10min.

4.5.5 Label a tube E_{mx} (pepsin enzyme, no protein control): 80 μL distilled water were added to 1.52 mL SGF plus pepsin, and then were mixed.

4.5.5.1 Immediately 200 μL into the E0 tube were removed, mixed and heated at 85°C for 10min.

4.4.5.2 After 60 minutes at 37°C water bath, 200 µL into the E60 tube were removed, mixed and heated at 85°C for 10min.

4.5.6 Label a tube D_{mx} (digestion mixture): 80 µL out of tube P was added to 1.52 mL SGF plus pepsin and mixed, then placed in 37°C water bath.

4.5.6.1 At 0.5, 2, 5, 10, 20, 30, 60 min internals, 200 µL of digestion mixture were withdrawn into D0.5, D2, D5, D10, D20, D30, D60 quenching tubes. (e.g. D0.5 at 30 sec., D2 at 2 min), each sample tube was heated to 85°C for 10 min.

4.5.7 P_0: 190 µL of SGF plus pepsin was added, quick heated at 85°C, then 10 µL out of tube P was added. Solution was vortexed and then heated at 85°C for 10 min.

4.6 SDS-PAGE Gel. All samples on any one gel were from a single digestion experiment. Novex 10-20% tris-glycine gels were used with SDS-PAGE buffer.

4.6.1 10 µL of each sample tube was loaded per well, containing 1.17 mg of starting LegHb per well except in wells for $P_{1/10}$ tube).

4.6.2 5 µL of pre-stained precision plus protein™ Dual Xtra Standards molecular weight marker proteins were loaded in the outer two wells.

4.6.3 Electrophoresis was accomplished at a constant 125 vdc.

4.6.4 ! Gels for staining were fixed in 10% acetic acid, 50% methanol for 15 min, then stained for a minimum of 6 hours in EZBlue as detailed by Sigma, then destained for at least 30min in water.

4.7 Image Analysis. The destained gels were visualized in a Gel Logic 440 Image Station under white light trans-illumination. The image was captured and the image intensity adjusted to optimum background and band intensities. The raw image was saved as an archival file.

4.7.1 The molecular weight of the BSA, ovalbumin, LegHb and any resulting degradation band that was not in the pepsin only lane was noted.

4.7.2 The 10% control band (P1/10) was used as the standard for comparison of all digested samples on a given gel.

4.7.3 ! The first time point the digested band appeared to be less than the10% concentrated sample was used to estimate the time to achieve 90% digestion.

4.8 Proteomic LC-MSMS Identity of Faint 22 kDa Band. Since a faint, diffuse band was visible only in digestion samples of LegHb, at 22 kDa, there was an attempt to identify the protein(s) at the Protein Core facility at the University of Nebraska. Residual samples were electrophoretically separated on a duplicate gel and the bands stained at 22 kDa were excised, digested by trypsin and analyzed by LC-MSMS using a LTQ Pro Velos Dual Pressure Linear Ion Trap mass spectrometer (ThermoScientific). The analyst constructed a genome complete soybean specific library using the NCBI soybean dataset. Human keratin sequences were masked and resulting peptides were screened against soybean, *E. coli* and porcine pepsin sequences from NCBI.

5. RESULTS & DISCUSSION

5.1 Limit of Detection. The stained gel of the dilution series of LegHb (Figure 1) demonstrated a clear pattern of reduced intensity of stained bands with each step in the dilution series. The minimum amount of protein that was detected was 15 µg for LegH. The concentration of LegH was labeled ~4 mg/ml. Based on these data, the limit of detection was approximately 1.25% (15 µg of LegHb) of leghemoglobin 100% loading (1.19 g of LegH) used in the digestion samples. This level of sensitivity was clearly sufficient to detect 10% residual of either hemoglobin in the digest.

5.2 Pepsin Activity. The certified activity of the lot of pepsin from Worthington used in this study was 2,730 units per mg of solid.

5.3 Control Substance Digestion Results. Stained gels of digestion tests of control substance BSA and ovalbumin (Figure 2 and 3) demonstrated that BSA was digested rapidly within the SGF plus pepsin test system and that ovalbumin was stable with more than 10% stainable full-length protein band remaining at 60 minutes. This result was consistent with results from previous test (Ofori-Anti, A.O. 2008), which showed the usability and reliability of this SGF plus pepsin test system.

5.4 Test Protein Digestion Results. A representative stained gel of digestion experiments of LegHb at pH 2.0 (Figure 4) demonstrated that the LegHb protein was stable in acid alone for 60 minutes (lane 3), but rapidly digested by pepsin in 0.5 minutes (lane 5) to below the visible band intensity of the quenched pepsin 10% hemoglobin control (P1/10 control in lane 14). There were other very faint lower molecular weight bands (~27 kD and ~ 30 kD) visible in all lanes that contained pepsin, including lanes 12 and 13, the pepsin only controls, and these were considered to be contaminants from the pepsin preparation.

5.5.1 ! In lane 5, there was a faint band at ~13 kDa, that we considered to be an intermediate fragment of LegHb. This was rapidly digested as it was invisible at time 2 minutes (lane 6).

5.5.2 A stained diffuse 22 kDa protein band was apparent that was analyzed by the protein core facility at UNL. Following trypsin digestion and LC-MSMS no fragments of the target test protein, LegHb were identified. Instead there was a positive peptide match to a 22 kDa protein of soybean, C6T9A3, a protein of unknown function. The protein may represent a culture media carry-over protein from the *E. coli* culture to produce the LegHb protein, although that cannot be absolutely proven, or an artifact. The sequence coverage representing the identified peptide was only 14%. In addition two short peptides from pepsin were identified, which is to be expected from auto-digestion of the pepsin digestion assay. In addition, peptides from three *E. coli* proteins were identified as minor components (thioredoxin, helicase and D-alanine ligase). There is no evidence of a stable fragment of LegHb digestion.

6. CONCLUSIONS

The results of this study demonstrated that the LegHb protein was rapidly digested after incubation in SGF plus pepsin at 37°C, with more than 90% digested within 0.5 minutes based on EZBlue. Pepsin-stable fragments of the LegHb were not identified and it seems there is no evidence that should cause concern related to food allergy based on these results.

7. REFERENCES

Astwood, J. D., Leach, J. N., and Fuchs, R. L. (1996). Stability of food allergens to digestion *in vitro*. *Nature Biotech.* **14**: 1269-1273.

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.

Codex Alimentarius (2003). Codex 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. Rome, *Codex Alimentarius* Commission, 2003, pp 47-60.

Collins, J.F., Fine, R. (1981). The enzymatic digestion of elastin at acidic pH. *Biochim. Biophys. Acta* **657:** 295-303.

Crevieu-Gabriel, I, Gomez, J., Caffin, J.P., Carre, B. (1999). Comparison of pig and chicken pepsins for protein hydrolysis. *Reprod. Nutr. Devel*. **39:** 443-454.

del Val, G., Yee, B.C., Lozano, R.M., Buchanan, B.B., Ermel, R.W., Lee, Y.M., Frick, O.L. (1999). Thioredoxin treatment increases digestibility and lowers allergenicity of milk. *J. Allergy Clin. Immunol*. **103**: 690-697.

FAO/WHO (2001). Evaluation of the allergenicity of genetically modified foods. Report of a Joint FAO/WHO Expert Consultation on the Allergenicity of Foods Derived from Biotechnology. January 22-25, 2001. Rome, Italy.

Fu, T.J., Abbott, U.R., Hatzos, C. (2002). Digestibility of food allergens and non-allergenic proteins in simulated gastric fluid and simulated intestinal fluid-a comparative study. *J. Agric. Food Chem*. **50**: 7154-7160.

Ofori-Anti, A.O., Ariyarathna, H., Chen, L., Lee, H.L., Pramod, S.N., Goodman, R.E. (2008). Establishing objective detection limits for the pepsin digestion assay used in the assessment of genetically modified foods. *Reg. Toxicol. Pharmacol.* **52**:94-103.

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., Landry, T.D., MacIntosh, S.C., Rice, E.A., Privalle, L.S., Steiner, H.Y., Teshima, R., van Ree, R., Woolhiser, M., Zwadny, J. (2004). A multi-laboratory evaluation of a common in vitro pepsin digestion assay protocol used in assessing the safety of novel proteins. *Regul. Toxicol. Pharmacol*. **39**: 87-98.

Yagami, T., Haishima, Y., Nakamura, A., Osuna, H., Ikezawa, Z. (2000). Digestibility of allergens extracted from natural rubber latex and vegetable foods. *J. Allergy Clin. Immunol*. **106**: 752-762.

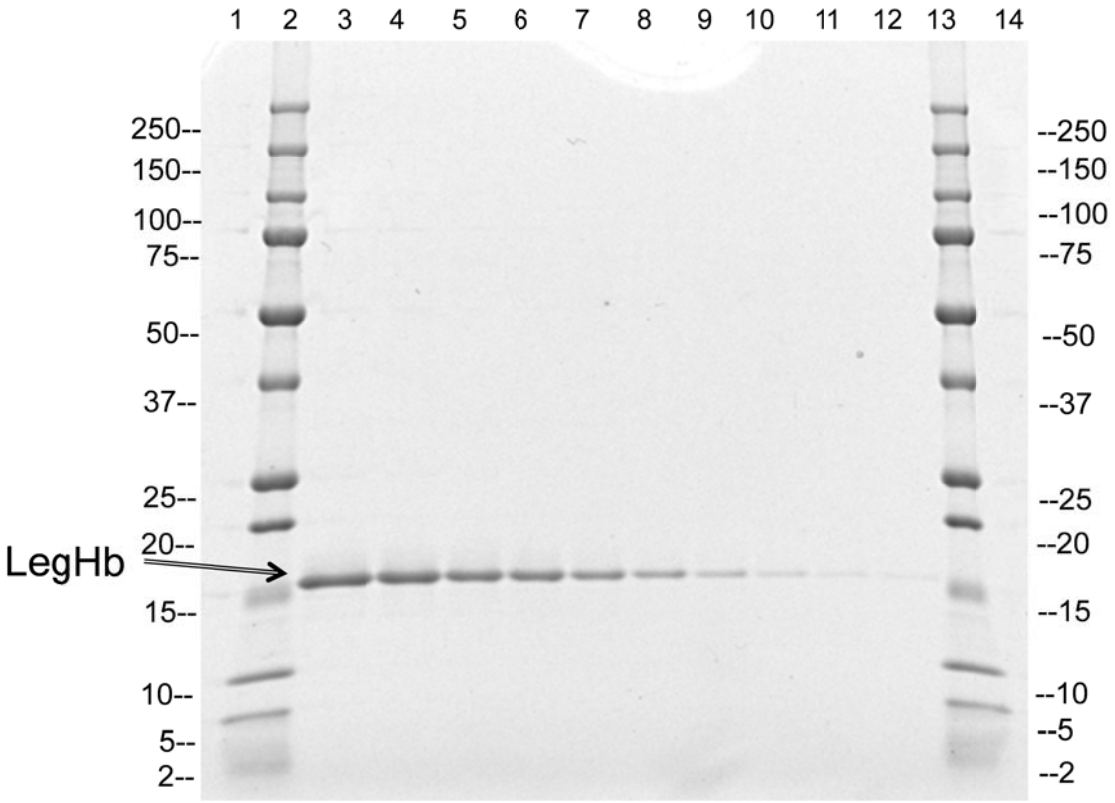


Figure 1. EZBlue Stained SDS-PAGE Gel Showing the Serial Dilution of Purified *E. coli* Produced LegHb. Proteins were separated by SDS-PAGE using a 10→20% polyacrylamide gradient in a glycine buffered gel. Proteins were detected by staining with EZBlue stain.

Lane	Description	Concentration (ug/ml)
1	Blank	na
2	Molecular weight marker	na
3	100% LegHb	139.0
4	75% LegHb	104.3
5	50% LegHb	69.5
6	40% LegHb	55.6
7	30% LegHb	41.7
8	20% LegHb	27.8
9	10% LegHb	13.9
10	5% LegHb	7.0
11	2.5% LegHb	3.5
12	1.25% LegHb	1.7
13	Molecular weight marker	na
14	Blank	na

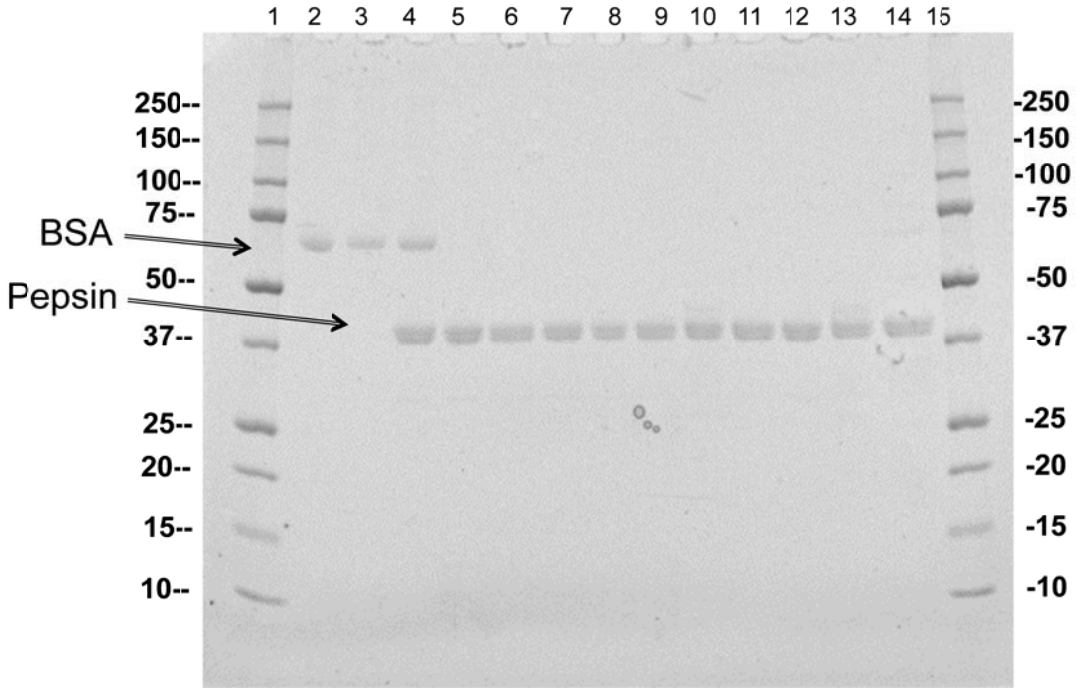


Figure 2. Representative EZBlue Stained SDS-PAGE Gel Showing the Digestion of BSA in Simulated Gastric Fluid. (pH 2.0) Proteins were separated by SDS-PAGE using a 10→20% polyacrylamide gradient in a glycine buffered gel. BSA was loaded 1.47 µg per lane based on pre-digestion concentration (pH 2.0). Proteins were detected by staining with EZBlue stain.

Lane	Description	Incubation time	
1	Molecular weight marker	na	
2	BSA without pepsin (P0)	0	min
3	BSA without pepsin (P60)	60	min
4	BSA in SGF, (D0)	0	min
5	BSA in SGF, (D0.5)	0.5	min
6	BSA in SGF, (D2)	2	min
7	BSA in SGF, (D5)	5	min
8	BSA in SGF, (D10)	10	min
9	BSA in SGF, (D20)	20	min
10	BSA in SGF, (D30)	30	min
11	BSA in SGF, (D60)	60	min
12	Experimental control pepsin (E0)	0	min
13	Experimental control pepsin (E60)	60	min
14	10% HPPD with quenched pepsin (P1/10)	0	min
15	Molecular weight Marker	na	

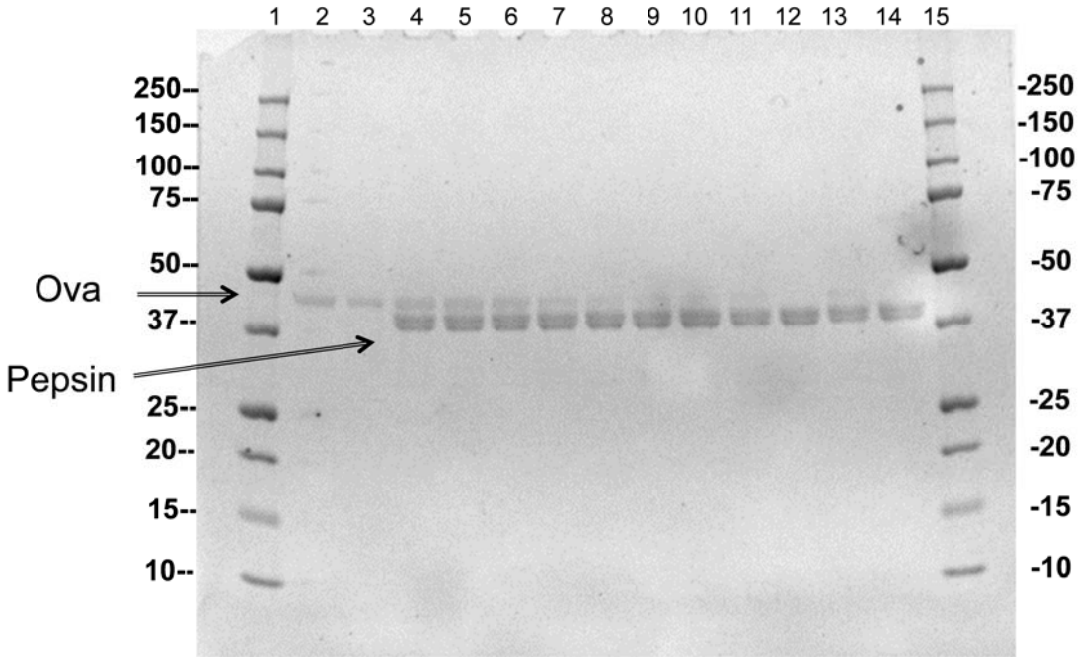


Figure 3. Representative EZBlue Stained SDS-PAGE Gel Showing the Digestion of Ovalbumin in Simulated Gastric Fluid. (pH 2.0) Proteins were separated by SDS-PAGE using a 10→20% polyacrylamide gradient in a glycine buffered gel. Ova was loaded 1.47 µg per lane based on pre-digestion concentration (pH 2.0). Proteins were detected by staining with EZBlue stain.

Lane	Description	Incubation time	
1	Molecular weight marker	na	
2	Experimental control without pepsin (P0)	0	min
3	Experimental control without pepsin (P60)	60	min
4	Ovalbumin in SGF, (D0)	0	min
5	Ovalbumin in SGF, (D0.5)	0.5	min
6	Ovalbumin in SGF, (D2)	2	min
7	Ovalbumin in SGF, (D5)	5	min
8	Ovalbumin in SGF, (D10)	10	min
9	Ovalbumin in SGF, (D20)	20	min
10	Ovalbumin in SGF, (D30)	30	min
11	Ovalbumin in SGF, (D60)	60	min
12	Experimental control pepsin (E0)	0	min
13	10% HPPD with quenched pepsin (P1/10)	0	min
14	Experimental control pepsin (E60)	60	min
15	Molecular weight Marker	na	

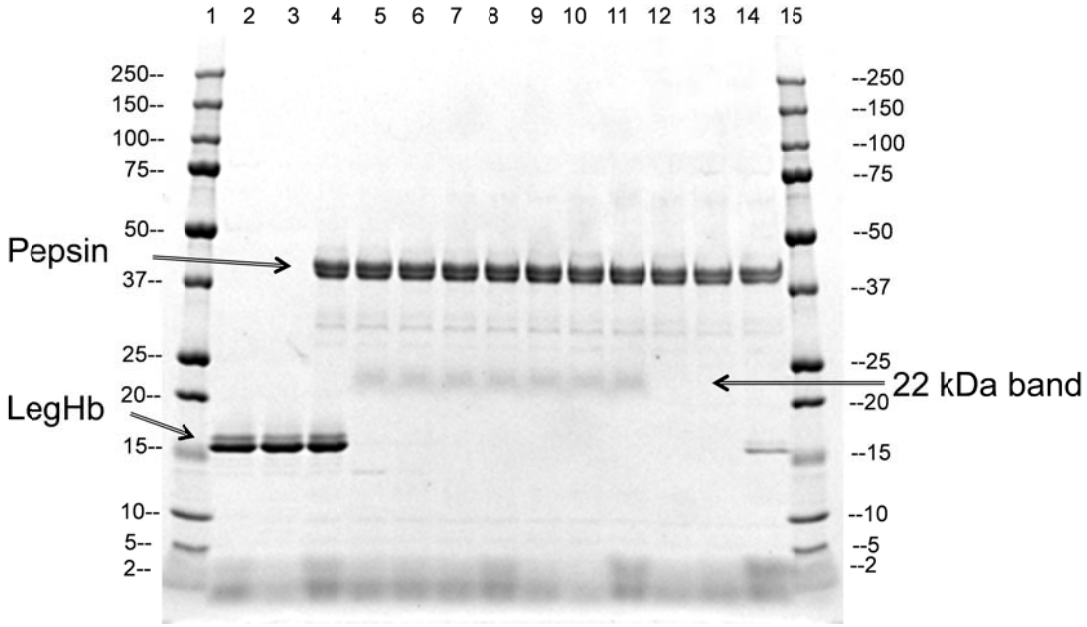


Figure 4. Representative EZBlue Stained SDS-PAGE Gel Showing the Digestion of LegHb in Simulated Gastric Fluid. (pH 2.0) Proteins were separated by SDS-PAGE using a 10→20% polyacrylamide gradient in a glycine buffered gel. LegHb was loaded 1.39 µg per lane based on pre-digestion concentration (pH 2.0). Faint 22 kDa band was tested and is primarily pepsin fragments

Lane	Description	Incubation time	
1	Molecular weight marker	na	
2	Experimental control without pepsin (P0)	0	min
3	Experimental control without pepsin (P60)	60	min
4	LegHb in SGF, (D0)	0	min
5	LegHb in SGF, (D0.5)	0.5	min
6	LegHb in SGF, (D2)	2	min
7	LegHb in SGF, (D5)	5	min
8	LegHb in SGF, (D10)	10	min
9	LegHb in SGF, (D20)	20	min
10	LegHb in SGF, (D30)	30	min
11	LegHb in SGF, (D60)	60	min
12	Experimental control pepsin (E0)	0	min
13	Experimental control pepsin (E60)	60	min
14	10% HPPD with quenched pepsin (P1/10)	0	min
15	Molecular weight Marker	na	

Annex 5
Batch analyses for three independent lots of RUBIA

Impossible Foods performed three RUBIA production runs in April-May 2014: PP-PGM2-14-120, PP-PGM2-14-125, and PP-PGM2-14-127. For each production run, the chemical composition of liquid RUBIA was determined by Silliker Inc. (Salida, CA). Impossible Foods' QA/QC Department tested RUBIA from each batch for total aerobic plate counts AOAC OMA 990.12 as well as total protein concentration, leghemoglobin purity, and heme concentration. AEMTEX Laboratories (Fremont, CA) tested RUBIA from each batch for *Salmonella* AOAC OMA 2011.03, *Listeria monocytogenes* AOAC OMA 2010.02, and *E. coli* O157:H7 AOAC RI 020801. All three batches fall within the specifications outlined in the typical analysis table (Section I.i.). Additionally, Impossible Foods tested RUBIA from each run to ensure the absence of 1) the production organism, 2) transformable recombinant DNA, and 3) antibiotic resistance genes.

I. RUBIA composition

i. Specifications of RUBIA and summary of liquid RUBIA % composition from three batch analyses.

	RUBIA Specifications		Batch Analysis (Frozen Liquid) (% w/w)		
	Concentration (Frozen Liquid) (% w/w)	**Concentration (Dry Powder) (% w/w)**	**PP-PGM2-14-120**	**PP-PGM2-14-125**	**PP-PGM2-14-127**
Protein	10	91	9.71	9.76	10.03
Leghemoglobin*	8	73	7.90	8.09	8.02
Ash	<1	<9	0.84	0.84	0.85
Fat	<0.1	<1	0.03	0.07	0.07
Carbohydrate	<0.1	<1	<0.1	0.11	<0.1
Solids	11	100	10.24	10.78	10.51
Moisture	89	0	89.76	89.22	89.49

*Impossible foods determined leghemoglobin concentration by multiplying total protein (% w/w) by leghemolgobin purity as determined by gel densitometry (Section I. vi.)

	RUBIA Specifications		Batch Analysis (Frozen Liquid) (ppm w/w)		
	Concentration (Frozen LIquid) (ppm w/w)	**Concentration (Dry Powder) (ppm w/w)**	**PP-PGM2-14-120**	**PP-PGM2-14-125**	**PP-PGM2-14-127**
Lead	<0.01	<0.1	<0.01*	<0.01*	<0.01*
Arsenic	<0.01	<0.1	<0.01*	<0.01*	<0.01*
Mercury	<0.005	<0.05	<0.005*	<0.005*	<0.005*
Cadmium	<0.1	<1	<0.1**	<0.1**	<0.1**

*Below Silliker level of detection
**Below the specification for RUBIA

	RUBIA Specifications		Batch Analysis (Frozen Liquid)		
	Concentration (Frozen Liquid)	**Concentration (Dry Powder)**	**PP-PGM2-14-120**	**PP-PGM2-14-125**	**PP-PGM2-14-127**
Aerobic plate count[1] (CFU/g)	<10^4	<10^4	<10	<10	<10
E. coli O157H7[2]	Absent by test	Absent by test	Absent by test	Absent by test	Absent by test
Salmonella spp[3]	Absent by test	Absent by test	Absent by test	Absent by test	Absent by test
Listeria monocytogenes[4]	Absent by test	Absent by test	Absent by test	Absent by test	Absent by test

[1] AOAC OMA 990.12
[2] AOAC RI 020801
[3] AOAC OMA 2011.03
[4] AOAC OMA 2010.02

ii. Total protein concentration of RUBIA was measured using a Pierce 660 absorbance assay (Pierce) and bovine serum albumin as a reference.

	PP-PGM2-14-120	**PP-PGM2-14-125**	**PP-PGM2-14-127**
Total protein concentration (mg/ml)	100.0	100.0	99.5

iii. Leghemoglobin purity was determined using Image Lab™ Software (Bio-Rad). Lanes were detected manually. Bands were detected automatically using high sensitivity followed by manual adjustment. The lane background was subtracted using a rolling ball disc size of 8.5.

	PP-PGM2-14-120	**PP-PGM2-14-125**	**PP-PGM2-14-127**
Leghemoglobin Purity (% total protein)	81.4	82.9	80.0

iv. Heme concentration in RUBIA was determined by UV-vis spectral analysis. Heme-bound globin proteins have a characteristic absorbance at 405-415 nm, which is referred to as the soret peak, as well as two smaller peaks in the 500 nm region. The ratio of the absorbances at 411 nm (soret peak) and 280 nm (total protein) are consistent with an 80% pure leghemoglobin solution that is 100% heme-bound.

	PP-PGM2-14-120	**PP-PGM2-14-125**	**PP-PGM2-14-127**
411abs/280abs	2.68	2.75	2.77

II. Absence of production organism in RUBIA

i. 0.5 ml of RUBIA was plated onto YPD+NAT (50 ug/ml nourseothricin) and grown for 72 hrs at 30 degrees Celsius.

ctive plating on YPD+N

	PP-PGM2-14-120	**PP-PGM2-14-125**	**PP-PGM2-14-127**
Total aerobic count on YPD+NAT (CFU/g)	0	0	2*

*NAT-resistant colonies were examined microscopically and lacked *P. pastoris* cell morphology. Additionally, colony PCR with MXY022-specific primers confirmed that the colony did not contain genomic DNA of the production organism.

III. Absence of antibiotic resistance genes and transformable DNA in RUBIA

i. Quantitative PCR Analysis: A standard phenol chloroform extraction was performed on 0.1g RUBIA to isolate any DNA that might be present in the sample. Quantitative PCR (qPCR) was performed on the extractions using two primer pairs targeting the AMP and NAT antibiotic resistance genes *AMPr* and *NATr*, respectively. The pJAN legH expression vector was used as a reference to generate a standard curve for each primer pair.

	PP-PGM2-14-120	PP-PGM2-14-125	PP-PGM2-14-127	Blank
Amplified DNA (pg) with NATr-specific primers	<0.001	<0.001	<0.001	<0.001
Amplified DNA (pg) with AMPr-specific primers	<0.001	<0.001	<0.001	<0.001

ii. Transformation Analysis: To test for the presence of transformable DNA, RUBIA was transformed into chemically competent *E. coli* cells (XL10 Gold Ultra Competent Cells, Invitrogen) using standard procedures. The pJAN legH expression vector was used as a positive control. Transformations were plated on YPD+NAT (50 ug/ml nourseothricin) and LB+AMP (100 ug/ml ampicillin) and grown at 37 degrees Celsius for 20 hrs.

	PP-PGM2-14-120	PP-PGM2-14-125	PP-PGM2-14-127	10 ng pJAN LegH Plasmid + PP-PGM2-14-106	10 ng pJAN LegH Plasmid
Total aerobic count on LB+AMP CFU/g	0	0	0	>500	>500
Total aerobic count on YPD+NAT CFU/g	0	0	0	>500	>500

Annex 6
Detection of non-symbiotic hemoglobin protein in mung bean sprouts

Plant non-symbiotic hemoglobins and leghemoglobins are structurally very similar and contain an identical heme B cofactor (Annex 1). Non-symbiotic plant hemoglobins have a history of safe human consumption though sprouted soybean and barley, as well as rice embryos (Anderson C. R., 1996) (Durnin, 1998) (Lira-Ruan, Ruiz-Kubli, & Arredondo-Peter, 2011). To test whether non-symbiotic hemoglobin protein is expressed in the widely consumed mung bean sprouts, Impossible Foods developed an ion exchange purification method to isolate non-symbiotic hemoglobin from mung bean sprouts (Sprouts were purchased at Whole Foods, December 2012). The isolated material was further separated by size using SDS-PAGE. Protein bands between 10-20 kD were excised and sent to MS Bioworks (Ann Arbor, MI) for nano-LC/MS/MS analysis. Three tryptic non-symbiotic hemoglobin peptides were detected in the sample. This positive identification indicates that mung bean sprouts that are readily consumed in the human diet contain non-symbiotic hemoglobin protein.

Anderson, C. R., Jensen, E. O., Leewellyn, D. J., Dennis, E. S., & and Peacock, W. J. (1996, June). A new hemoglobin gene from soybean: A role for hemoglobin in all plants. *Proc. Natl. Acad. Sci. USA, 93*, 5682-5687.

Durnin, D. C. (1998). Haemoglobin expression in germinating barley. *Seed Science Research*, 431-436.

Lira-Ruan, V., Ruiz-Kubli, M., & Arredondo-Peter, R. (2011). Expression of non-symbiotic hemoglobin 1 and 2 genes in rice (Oryza sativa) embryonic organs. *Communicative and Integrative Biology*, *4*(4), 457-458.

MS Bioworks
Biological Mass Spectrometry Solutions

3950 Varsity Drive
Ann Arbor, MI 48108
734-929-5083
www.msbioworks.com
info@msbioworks.com

Project Report (

Information

Client:	Marija Vrljic
Institute:	Sand Hill Foods
MSB Project Number:	MSB-2921
Date submitted:	11/09/12
Date completed:	12/04/12

Samples

Client identifier	MSB identifier
A 1 µg	14219
B 0.1 µg	14220
C 0.01 µg	14221
D UK	14222

Objective

Identification of proteins in submitted gel band and evaluation of detection levels for assay development.

Experimental Methods

Sample Preparation

In gel digestion was performed using a ProGest robot (DigiLab) with the following protocol:

- Washed with 25mM ammonium bicarbonate followed by acetonitrile.
- Reduced with 10mM dithiothreitol at 60℃ followed by alkylation with 50mM iodoacetamide at RT.
- Digested with trypsin (Promega) at 37℃ for 4h.
- Quenched with formic acid and the supernatant was analyzed directly without further processing.

Mass Spectrometry

Each gel digest was analyzed by nano LC/MS/MS with a Waters NanoAcquity HPLC system interfaced to a ThermoFisher LTQ Orbitrap Velos. Peptides were loaded on a trapping column and eluted over a 75µm analytical column at 350nL/min; both columns were packed with Jupiter Proteo resin (Phenomenex). The mass spectrometer was operated in data-dependent mode,

with MS performed in the Orbitrap at 60,000 FWHM resolution and MS/MS performed in the LTQ. The fifteen most abundant ions were selected for MS/MS.

Data Processing

Data were searched using a local copy of Mascot with the following parameters:

Enzyme: Trypsin
Database: Custom including the target sequence* (concatenated forward and reverse plus common contaminants)
Fixed modification: Carbamidomethyl (C)
Variable modifications: Oxidation (M), Acetyl (N-term), Pyro-Glu (N-term Q), Deamidation (N,Q)
Mass values: Monoisotopic
Peptide Mass Tolerance: 10 ppm
Fragment Mass Tolerance: 0.015 Da
Max Missed Cleavages: 2

*> Non-symbiotic hemoglobin [Vigna radiata] Moong bean

MTTTLERGFTEEQEALVVKSWNVMKKNSGELGLKFFLKIFEIAPSAQKLFSFLRDSTVPLEQNP
KLKPHAVSVFVMTCDSAVQLRKAGKVTVRESNLKKLGATHFRTGVANEHFEVTKFALLETIKEA
VPEMWSPAMKNAWGEAYDQLVDAIKYEMKPPSS

Mascot DAT files were parsed into the Scaffold software for validation, filtering and to create a non-redundant list per sample. The data were filtered using a minimum protein value of 90%, a minimum peptide value of 50% (Prophet scores) and requiring at least two unique peptides per protein.

Results

A summary of the data are included in the table below and greater details regarding the identified proteins can be found in the Excel work book MSB-2921 Sand Hill Foods 120412.xls.

The Excel file contains the **Protein Report** which has the full list of proteins identified (including known contaminants) and their molecular weight and spectral counts (SpC). A summary is provided here for Non-symbiotic hemoglobin [Vigna radiata] Moong bean:

Identified Proteins	Accession Number	Molecular Weight	SpC 14219	SpC 14220	SpC 14221	SpC 14222
hemoglobin Vigna Radiata Moong Bean	Non-symbiotic	18 kDa	346	279	158	3

Sequence coverage details from sample 14222 are below. Three tryptic peptides were detected all of which are suitable candidates for assay development.

Non-symbiotic (100%), 18,138.2 Da
hemoglobin Vigna Radiata Moong Bean
3 unique peptides, 3 unique spectra, 3 total spectra, 30/161 amino acids (19% coverage)

MTTTLERGFT EEQEALVVKS WNVMKKNSGE LGLKFFLKIF EIAPSAQKLF SFLRDSTVPL
EQNPKLKPHA VSVFVMTCDS AVQLRKAGKV TVRESNLKKL GATHFRTGVA NEHFEVTKFA
LLETIKEAVP EMWSPAMKNA WG AYDQLVD AIKYEMKPPS S

A Scaffold file is also provided with this report; the Scaffold file can be downloaded from our FTP site using the details provide below.

FTP Details:

FTP://75.144.89.5

Username: Sand Hill Foods
Password: 3gQBz1JT

SUBMISSION END

From: Yingling, Gary L.
To: Bonnette, Richard
Subject: RE: Submission to FDA"s GRAS Notification Program (soybean leghemoglobin from Pichia pastoris)
Date: Wednesday, September 24, 2014 1:54:20 PM

Dear Mr. Bonnette: I have spoken with the officials at Impossible Foods, Inc and they have informed me that the confidential markings on the two reports which are attachments in the GRAS Notice were marked that way for submission to Impossible Foods or a related firm and the documents are no longer confidential. There are no confidential documents in the Impossible Food. Inc GRAS Notice submission. gary

Gary L. Yingling
Morgan, Lewis & Bockius LLP
1111 Pennsylvania Avenue, NW | Washington, DC 20004-2541
Direct: +1.202.739.5610 | Main: +1.202.739.3000 | Fax: +1.202.739.3001
gyingling@morganlewis.com | www.morganlewis.com

From: Bonnette, Richard [mailto:Richard.Bonnette@fda.hhs.gov]
Sent: Tuesday, September 23, 2014 12:28 PM
To: Yingling, Gary L.
Subject: Submission to FDA's GRAS Notification Program (soybean leghemoglobin from Pichia pastoris)

Mr. Yingling,
Regarding your submission to the FDA GRAS Notification Program on behalf of Impossible Foods, Inc. for uses of soybean leghemoglobin from *Pichia pastoris,* a question has come up regarding the "confidential" indications on pages 46-70 (FARRP Bioinformatics Report pages 1-25) and pages 98-100 (MS Bioworks report).

I suspect that these pages may have been marked as confidential as matter of process by the laboratories and researchers who prepared these reports and included in the submission as an oversight rather than a claim of confidentiality by Impossible Foods, Inc. Can you confirm whether or not the notifier intends to claim the information as confidential on those pages?

Best regards and thanks,
Richard Bonnette

Richard E. Bonnette, M.S.
Consumer Safety Officer
Division of Biotechnology and GRAS Notice Review
Office of Food Additive Safety
U.S. FDA, Center for Food Safety and Applied Nutrition

(240)402-1235
Richard.Bonnette@fda.hhs.gov

Mailing address:
5100 Paint Branch Parkway, HFS-255

College Park, MD 20740

DISCLAIMER
This e-mail message is intended only for the personal use of the recipient(s) named above. This message may be an attorney-client communication and as such privileged and confidential and/or it may include attorney work product. If you are not an intended recipient, you may not review, copy or distribute this message. If you have received this communication in error, please notify us immediately by e-mail and delete the original message.