

Evaluating potential risks of food allergy and toxicity of soy leghemoglobin expressed in *Pichia pastoris*

Yuan Jin¹, Xiaoyun He², Kwame Andoh-Kumi¹, Rachel Z. Fraser³, Mei Lu¹ and Richard E. Goodman^{1*}

¹ Food Allergy Research and Resource Program, Dept. of Food Science & Technology, University of Nebraska-Lincoln, Lincoln 68508, NE, USA

² College of Food Science and Nutritional Engineering, China Agricultural University, No. 17 Tsinghua Donglu, Beijing, 100083, China

³ Impossible Foods Inc. 525 Chesapeake Dr. Redwood City, CA, 94063, USA

* Correspondence to: Richard E. Goodman, 283 Food Innovation Center, Lincoln, NE, 68588-6207 Phone: 402-472-0452

Email: rgoodman2@unl.edu

Key words: allergenicity, bioinformatics, toxicity, GMO, hemoglobin

Abbreviations: AA: amino acid; AOL: AllergenOnline database; FAO: Food and Agricultural Organization; HOSU: history of safe use; LegHb: Leghemoglobin; NCBI: National Center for Biotechnology Information; OVA: ovalbumin; SGF: simulated gastric fluid; ALDH: aldehyde dehydrogenase, G6PD: glucose 6 phosphate dehydrogenase, REP: recombinantly expressed protein.

Abstract:

Scope

The Soybean (*Glycine max*) *leghemoglobin c2* (LegHb) gene was introduced into *Pichia pastoris* yeast for sustainable production of a heme-carrying protein, for organoleptic use in plant-based meat.

The potential allergenicity and toxicity of LegHb and 17 *Pichia* host-proteins each representing ≥1% of total protein in production batches were evaluated by literature review, bioinformatics sequence comparisons to known allergens or toxins, and *in vitro* pepsin digestion.

Methods and results

Received: 17-04-2017; Revised: 09-08-2017; Accepted: 30-08-2017

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/mnfr.201700297](#).

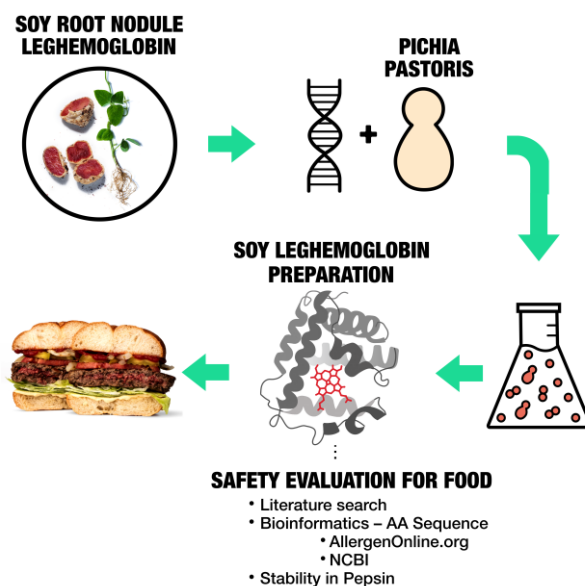
This article is protected by copyright. All rights reserved.

Literature searches found no evidence of allergenicity or toxicity for these proteins. There were no significant sequence matches of LegHb to known allergens or toxins. Eleven *Pichia* proteins had modest identity matches to minor environmental allergens and thirteen *Pichia* proteins had significant matches to proteins from toxic sources. Yet the matched allergens and toxins had similar matches to proteins from the commonly consumed yeast *Saccharomyces cerevisiae*, without evidence of food allergy or toxicity. The demonstrated history of safe use indicates additional tests for allergenicity and toxicity are not needed.. The LegHb and *Pichia sp.* proteins were rapidly digested by pepsin at pH 2.

Conclusion

These results demonstrate that foods containing recombinant soy LegHb produced in *Pichia sp.* are unlikely to present an unacceptable risk of allergenicity or toxicity to consumers.

Soy leghemoglobin protein (LegHb) is found in soy root nodules. The protein was produced in *Pichia* yeast for high-quality production and is used to improve the flavor profile of plant-based meat. Literature review, protein sequence analysis, and *in vitro* pepsin digestion demonstrate that foods containing LegHb produced in *Pichia* are unlikely to present a risk of allergenicity or toxicity.



1. Introduction

Hemoglobins are ubiquitous iron binding proteins in nature, present in bacteria, fungi, higher plants and animals {Everse, 2004 #24}. Consumption of these proteins serves as an efficient source of bioavailable iron, which is required for oxygen transport, respiration, and other metabolic functions {Abbaspour, 2014 #3}. Animal hemoglobins and myoglobins have long been widely consumed in the human diet through meat, poultry and fish products. Dietary exposure to plant hemoglobins is also common. Sprouted barley, widely used in the beverage industry (malted barley) and in the baking industry (malted barley flour), has been shown to express hemoglobin 2-3 days after imbibition {Duff, 1998 #21}. In legumes, four major leghemoglobin (LegHb) isoproteins are generally found in root nodules, where atmospheric N₂ is reduced to ammonia and assimilated in exchange for photosynthetically produced sugars {Appleby, 1980 #5; Fuchsman, 1979 #29}. Leghemoglobin is structurally similar to the widely consumed mammalian myoglobins {Hargrove, 1997 #150}. They share an identical heme cofactor (heme B), which binds oxygen with high affinity. These mammalian myoglobins share a common history of safe use (HOSU) in foods.

Impossible Foods Inc. (Redwood City, CA) discovered that heme proteins are critical to the flavor profile of animal-derived meat (US patent 14/797,006, allowed March 22, 2017). Therefore, to recreate the flavor profile of animal-derived meat in plant-based meat products, Impossible Foods synthesized the *leghemoglobin c2* gene from the soybean plant, inserted the gene into a yeast host, *Pichia pastoris* (now reclassified as *Komagataella phaffii*), and purified the LegHb protein expressed in the resulting *Pichia* strain to >65% purity. The LegHb protein, produced in this way, has been tested in plant-based meat products at concentrations up to 0.8%. The LegHb c2 gene from soybean (*Glycine max*) root nodules encodes a 145 amino acid (AA) protein sequence listed as Accession number P02236 in the UniProt protein database. Seventeen *P. pastoris* proteins, normally expressed by the yeast host, that remain as minor components in the highly enriched LegHb protein product, were identified by Impossible Foods and the UC Davis Genome center using proteomic analysis. Each *Pichia* protein is present at concentrations of less than 0.1% in the plant-based meat product.

The yeast *P. pastoris* is widely used as a host for recombinant protein expression {Spohner, 2015 #66}. *Pichia* belongs to the same family of yeast (*Saccharomycetaceae*) as several yeast genera widely used in food: *Saccharomyces*, *Torula*, *Yarrowia*, *Dekkera* and *Brettanomyces*. *Brettanomyces*, a yeast traditionally used in brewing Belgian beers, belongs to the same sub-family of yeast as *Pichia* - the *Pichiaceae*. Yeast extract (from *S. cerevisiae* and *Torula*) is frequently consumed in substantial quantities in human diets. *Pichia* is used for the GRAS-approved production of BD16449 *Phospholipase C*, an enzymatic processing aid for degumming vegetable oils (GRN 204). Additionally, *Pichia* is approved as an animal feed source (21 CFR §573.750) and has been used to produce a number of FDA-approved pharmaceuticals (<http://www.pichia.com/science-center/commercialized-products/>). However, *Pichia* is most commonly used for expression of secreted recombinant proteins since it produces recombinant proteins more efficiently than does *Saccharomyces*. At present, there are no reports of *Pichia* intracellular host proteins in human food. The potential for allergenicity and toxicity of the final LegHb product and the seventeen most prominent *Pichia* host proteins were evaluated for safety as food ingredients.

The primary concerns for food safety from any recombinantly expressed protein (REP) are whether the protein itself is allergenic, whether the protein is cross-reactive due to similarity to another protein, whether the protein is a toxin, or whether insertion of the gene alters the quantity of endogenous allergens if the host is commonly allergenic {Goodman, 2008 #34;Hammond, 2013 #35}. The CODEX Alimentarius Commission recommended a weight-of-evidence approach to judge whether the introduced novel protein has allergy risk or not{Ladics, 2008 #41;Codex, 2003 #16}. The approach involves a multi-step process, including review of the allergenicity, or safety of the gene source, sequence similarity comparison of the protein to known allergens, and testing the stability of the protein to digestion by pepsin. Understanding the abundance of the protein in the food products is relevant as well. Since many allergens and toxins have now been identified, a bioinformatic sequence comparison combined with a literature review serves as an important tool for assessing the potential risks {Codex, 2003 #16;Goodman, 2008 #34}. If a significant sequence

identity match is found, serum IgE tests using samples from subjects sensitized to the source may be needed for evaluation of the allergenic potential {Goodman, 2008 #32}. The potential toxicity assessment recommended by CODEX includes a review of toxins or toxicity of the source organism and a comparison of the protein AA sequence identity between the novel protein and known protein toxins {Mathesius, 2009 #47}. Tests to measure the stability of the protein in an *in vitro* gastric digestion model are also used to judge potential risks of potential allergenicity or toxicity of the protein {Codex, 2003 #16;Ladics, 2008 #41;Ladics, 2008 #41;Goodman, 2008 #34;Ladics, 2008 #41}.

Allergenic cross-reactivity is difficult to predict. Aalberse suggested that proteins with <50% identity in AA sequence over full-length with an allergen are unlikely to be cross-reactive {Aalberse, 2000 #2}. However, CODEX recommended using criteria for matches of >35% identity over 80 AA by FASTA as a conservative prediction for potential cross-reactivity. Although not stated in the guideline, the 80 AA-segment match is apparently based on the rationale that either natural or man-made genetic changes might transfer a structural motif that could include sequential or conformational IgE-binding epitopes capable of cross-linking IgE on receptors on sensitized mast cells or basophils that could induce allergic mediator release {Codex, 2003 #16;Gieras, 2011 #31}. Some regulatory authorities require a search comparison for 8 AA segment identity matches to allergens to identify possible cross-reactive targets {Codex, 2003 #16;Ladics, 2008 #41}.

Resistance to pepsin digestion was recommended by CODEX as a central part of the allergenicity assessment of recombinant proteins intended for human consumption based on earlier publication by Metcalfe et al. in 1996 {Metcalfe, 1996 #49}. The digestion assay is not meant to predict whether a given protein will always be digested in the stomach of human consumers, but the evaluation provides a simple *in vitro* correlation between resistance to pepsin digestion and allergenic potential [{Fu, 2002 #28;Ofori-Anti, 2008 #54;Bannon, 2002 #9}].

Evaluation of the safety of recombinant LegHb protein and 17 minor yeast proteins were evaluated by literature review, bioinformatic analysis for sequence similarity to proteins in the

AllergenOnline (AOL) and NCBI databases, as well as *in vitro* pepsin digestion by simulated gastric fluid (SGF) at two pepsin-protein ratios, 10:1 and 1:1 (U/μg).

2. Methods

2.1 Scientific literature review

Four public literature databases were used to search for publications indicating possible risk using combinations of keywords (soybean or *Glycine max* with hemoglobin and allergy, allergen, toxic or toxin as well as the 17 high *Pichia pastoris* proteins) to identify articles that might indicate potential allergenicity or toxicity of the protein or the host organism. These four databases were PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>), Web of Science™ Core Collection [v.5.24] (<http://apps.webofknowledge.com/>), Scopus (<https://www.scopus.com/>), and National Agricultural Library Catalog (AGRICOLA, <https://agricola.nal.usda.gov/>). The original searches with PubMed were conducted on 07 March 2016 and 01 July 2016 (leghemoglobin and the 17 *Pichia* proteins, respectively). These searches were repeated with PubMed, adding the Web of Science, Scopus and Agricola databases on 22 May, 2017. Publication abstracts were reviewed and identified publications searched for indications of potential indications of adverse health risks.

2.2 Sequence databases and bioinformatics search strategies

2.2.1 Sequence of LegHb and the 17 *Pichia* host proteins

Table 1 shows the full AA sequence of the LegHb protein. Figure 1 shows a Coomassie stained SDS-PAGE gel image of a production scale batch of Soy Leghemoglobin Preparation (PP-PGM2-15-320-101), provided by Impossible Foods Inc. Ten visible bands in addition to LegHb were evaluated by LC-MS/MS and found to contain 17 *Pichia* proteins, each band representing ≥1% of the total protein fraction of LegHb preparation based on densitometry. Each protein sequence was used to identify full-length matches to sequences in the NCBI database (Table S1). Those sequences were then used as query sequences in bioinformatics searches.

2.2.2 AllergenOnline (AOL) version 2016 (v16)

The AllergenOnline (AOL) database (<http://www.allergenonline.org/>) is a peer-reviewed allergen list and sequence database that is updated each year {Goodman, 2016 #33}. We used the AOL as the primary tool in searching for matches to allergenic sequences. Three comparison methods were used: an overall FASTA3 full-length search (using *E* scores of 10 and 1.0), a sliding 80 AA window FASTA search (*E* score of 10, identity > 35%), and an exact word search tool looking for 8 AA exact matches. The *E* score criteria limits have been established on the AllergenOnline database for more than five years based on extensive testing to limit irrelevant matches while identifying alignments that may cause cross-reactivity {Goodman, 2016 #33}.

2.2.3 BLASTP in NCBI Entrez protein database

The BLASTP search algorithm is available on the NCBI website (<http://www.ncbi.nlm.nih.gov/BLAST/>). Protein sequence entries in the Entrez search and retrieval system are compiled and maintained by the NCBI of the National Institutes of Health (U.S.A.). The database is updated or modified every few days. Sequences are annotated with publications or with notation of likely protein types; much of that is added by computer algorithm and some simply suggest minor similarities to a previously identified allergen type. Since AOL is updated once per year, the NCBI protein database was searched using keyword limits ("allergen", "toxin" and "toxic") to identify possible matches to proteins that may represent newly discovered allergens or toxins. The searches for matches between newly deposited proteins and the LegHb and the 17 host proteins, respectively, were performed on Feb 27 to 29, 2016 and May 27, 2016, using the BLASTP version 2.3.0. Searches were also performed for each sequence without any keyword to evaluate whether similar proteins exist in other organisms that might provide information of safe exposure to homologous proteins. The additional searches ensured that sequences, which might not yet have been entered into AOL v16, were included in the analysis {Siruguri, 2015 #65}. Additional bioinformatics searches with NCBI and BLASTP were used to compare sequence identity ranking with

proteins of risk or from known safe food sources. We evaluated the *E* score values, the length of the alignments, and the identity percentage of any identified match from the BLASTP results to judge the significance of any alignment.

2.3 *In vitro* pepsin digestibility study

The test system for this study was an *in vitro* digestion model using pepsin in SGF. The SGF preparation and digestion procedures were based on the methods described by Thomas et al. in 2004 {Thomas, 2004 #67}, as modified by Ofori-Anti {Ofori-Anti, 2008 #54}. Pepsin is optimally active between pH 1.2 and 2.0, but markedly less active at pH 3.5, and irreversibly denatured at pH 7.0 {Collins, 1981 #17;Créviu-Gabriel, 1999 #20}. Although the first pepsin digestion assay developed by Astwood recommended performing the digestion at pH 1.2 {Astwood, 1996 #7}, the FAO/WHO (2001) suggested using two pH conditions, pH 1.2 and pH 2.0 {FAO/WHO, 2001 #27}. 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 {Thomas, 2004 #67}. Moreover, we digested a number of proteins at both pH 1.2 and 2.0 and did not demonstrate significant differences {Ofori-Anti, 2008 #54}. Therefore in this study we only evaluated stability of the protein at pH 2.0.

2.3.1 Materials

Test and Control substances. The protein sample solution was provided by Impossible Foods Inc. from the LegHb production run PP-PGM2-15-320-101. The LegHb protein represented 66% of the total protein according to the certificate of analysis. The total protein concentration was determined to be 79.94 mg/ml using a GE 2D Quant kit (GE Healthcare, #80-6483-56), following the kit instructions. The PP-PGM2-15-320-101 is a representative batch of soy leghemoglobin protein preparation. The type and abundance of Pichia proteins is consistent from batch-to-batch. Bovine hemoglobin (Hb), (Sigma-Aldrich Co. LLC., St. Louis, MO, #H2625-25G), BSA, (Sigma-Aldrich

Co. LLC., St. Louis, MO #A9647-100G) and chicken ovalbumin (OVA), (Worthington Biochemicals, #3054) were used as control proteins in the digestion assay.

Pepsin A (Worthington Biochemicals, #3319) with a certificate of analysis of 2,810 activity units per mg solid was used in all assays. Novex® 10-20% tris-glycine polyacrylamide gels (Invitrogen) were used to separate digested materials. Precision Plus Protein™ dual xtra standards (BioRad) were used as the molecular weight standards. Tris-Glycine-SDS 10 x running buffer (Fisher Scientific), Coomassie Brilliant Blue R-250 staining solution, and Coomassie Brilliant Blue R-250 destaining solution (BioRad) were used for gel running, staining, and destaining, respectively.

2.3.2 Pepsin activity test

The activity of the pepsin was tested within 24 hours before each digestion assay. The assay was based on the method described by Worthington, as described by Ofori-Anti {Ofori-Anti, 2008 #54}. Pepsin A tested at 2826 units per mg solid right before the digestion assays, which was quite close to its labeled activity of 2810 U/mg.

2.3.3 Determination of system's limit of detection (LOD)

The concentration of the undigested protein in SDS-PAGE was considered to be 100%. The LOD of the test system must be lower than 10%, as the time for digestion is determined to be when the sample reaches 10% of starting material {Ofori-Anti, 2008 #54}. In order to establish the detection limit, a serial dilution of the sample solution was prepared with diluent and 5x reducing Laemmli buffer covering the range representing 200% total protein (2.96 µg) per well down to 2.5% (0.037 µg) per well. Samples were separated in an SDS-PAGE at a constant 150 voltage for 90 min, followed by staining with Coomassie blue. The LOD (Figure S1) was 5% of total protein at 100% loaded in the digestion samples. This level of sensitivity was clearly sufficient to detect 10% residual of leghemoglobin or any other protein in the digest.

2.3.4 Pepsin digestion resistance tests of LegHb and three control proteins

The tests were performed at two different pepsin-protein ratios, 10 units and 1 unit pepsin activity per 1 μ g substrate protein. The 10 units assay was described by Ofori-Anti {Ofori-Anti, 2008 #54}. The 1 unit assay was done to experimentally investigate the digestibility of the sample at a much lower pepsin-protein ratio, although this ratio has not been validated yet with many allergenic proteins. The “0 min” digestion sample and the undigested sample at 10% protein concentration were prepared by first quenching pepsin at 95°C for 10 min, then adding protein solution followed by another 10 min heating to denature the proteins.

3. Results

3.1 Literature search results

A literature review for the history of food safety of LegHb in four databases is summarized in Table S2. None of the publications implicated any allergenic effects of this protein. Searches with “Glycine max” AND “allergy” AND “heme” only yielded one reference from Pubmed database on impairment of carotenoid and flavonoid biosynthesis due to a mutation in Arabidopsis HY1 which is not relevant to the topic under review {Xie, 2012 #68}. A broader search of the literature demonstrates that soybean is reported to be a relatively common food allergen for young children, rarely for adults. Reactions are to proteins expressed in the seeds. There are no reports of allergy to the other parts of the plant except seed-pod proteins, referred by as Gly m 1 and Gly m 2 in the WHO/IUIS Allergen Nomenclature database (www.allergen.org/). No reports were identified for root-specific proteins. Since the LegHb is not present in the seed or seed-pods but only in root nodules, there is no reason to believe that people would be allergic to the protein and there is no need to perform serum IgE binding studies and there is no sensitive population to obtain relevant serum from for IgE binding studies.

Similarly, with the search filtered for keywords “toxin” or “toxicity”, the majority of the papers reported on the effect of heme-peroxidases or heme-oxygenase system and the adverse

effect due to changing of hydrogen peroxide level {Kim, 2011 #39}, which is not relevant to allergenicity or toxicity. Thus, based on literature search there is no reason to suspect the LegHb produced from *Glycine max* would raise a concern of allergy or toxicity.

A search for literature concerning the safe use of *Pichia pastoris* is summarized in Table S3. A review of the summary entries showed that they are related to the use of the yeast as a recombinant host for expressing a wide variety of proteins from various eukaryotic sources. In the Pubmed search, adding a third term “NOT recombinant” reduced the findings to five publications {Dumez, 2013 #22;Rodriguez-Perez, 2007 #59;Marazuela, 2007 #46;Yasuhara, 2001 #69;Kastenmüller, 2013 #38}. In the other three databases, the numbers were also reduced by adding keyword “NOT recombinant”. Examination of the results demonstrated no link to allergy related to endogenous proteins from *Pichia pastoris*. Searches with the revised name “*Komagataella phaffii*” without other terms greatly reduced the number of the results which were related to taxonomy or genomic cloning of the yeast species (synon. *Pichia pastoris*) or the use of this species as a recombinant host. Full evaluation of the potential risks of allergy of each of the identified 17 *P. pastoris* proteins was considered in searches in combination with “*Pichia pastoris*”. Review of the information failed to identify any study data to suggest possible allergenicity of the endogenous proteins. Thus, there is no published evidence that endogenous proteins from *Pichia pastoris* are allergens.

Searches using a combination of “*Pichia pastoris*” AND “toxin” yielded a large number of references in the four databases; most described expression of various eukaryotic genes to produce recombinant proteins. The list was reduced by excluding “recombinant”. One reference from Pubmed specifically examined toxin expression in the host, testing various *Pichia pastoris* (*Komagataella phaffii*) strains for toxins that are active against pathogens or other yeasts {Banerjee, 2000 #8}. They did not detect any evidence of toxic activity in the 14 tested strains. Another study published in identified in Scopus evaluated the Kpkt hertologous protein produced in *Pichia pastoris* {Chessa, 2017 #14}. This Kpkt protein occurs naturally in a different genus with similar species

name, *Tetrapisispora phaffii*. The Kpkt protein is active in reducing microbial spoilage in wine-production [Comitini, 2009 #18]. No other publications were identified that related to toxicity except those from heterologously expressed transgenes from other organisms. Thus, no evidence was found indicating that *Pichia pastoris* endogenous toxic proteins.

3.2 Bioinformatics search results of LegHb

Searches with the LegHb AA sequence against AOL v16 did not identify any significant alignment with an allergen. Full length alignments from FASTA3 searches with *E* scores less than 1.0 are shown in Table S4. Only one type of protein, the globin (insect hemoglobin) from *Chironomus thummi thummi*, aligned with <27% identity, which is below a level likely to indicate cross-reactivity [Aalberse, 2000 #2]. Ten additional alignments to taxonomically diverse sources (grass, house dust mite, potato and cockroach) were identified with *E* scores between 2.2 and 8 and are not shown. Neither the sliding 80-mer window searches nor eight-contiguous AA search resulted in any matches. Thus, it is unlikely that the LegHb protein is sufficiently similar to any known allergen to present a risk of cross-reactivity. The lack of sequence identity to known allergens as well as lack of allergy to the source (soybean root nodules), removes the normal requirement for serum IgE testing even though the protein is from soybean, a relatively commonly allergenic source.

The full-length sequence of LegHb was compared to all sequences in the NCBI-Entrez database without any keyword limit (data not shown). LegHb is highly similar and closely related to other plant hemoglobin proteins that bind oxygen, predominantly in other legumes. The protein buffers the oxygen concentration to enable symbiotic microbes in root nodules to fix nitrogen. LegHb is also approximately 26% identical to some chordate hemoglobins. The results indicate the high similarity of leghemoglobin to common plant and animal proteins without obvious indications of risks of allergy. BLASTP searches filtered with keywords “allergen” or “toxic” failed to identify matches, while BLASTP with keywords “allergy” or “toxin” produced low identity matches to sequences that are unlikely to be homologous due to low identities or short alignments (Table S5).

3.3 Bioinformatics search results of 17 *Pichia* proteins

Each of the 17 *Pichia* proteins were evaluated for potential matches to allergens in the AOL v16 database. The results are summarized in Table 2. Each protein was also compared to the NCBI-Entrez database without any keyword selection, and with keywords “allergen”, “toxin” or “toxic”. Summaries for each protein are shown in supplemental Tables S6 – S22. Overall, these bioinformatics search results showed little risk of food allergy or toxicity.

All of the 17 *Pichia pastoris* proteins are enzymes or an enzyme inhibitor (protein #10) that are ubiquitous in nature. Therefore, a search of the NCBI database for sequences related to each of the 17 proteins, using BLASTP without keyword limits identified high scoring alignments with related proteins from many molds and yeasts. In all cases, these included *Saccharomyces cerevisiae* and *Saccharomyces bayanus*, which are commonly used in making wine, bread, and beer and *Saccharomyces boulardii* which is widely used as a probiotic {Moyad, 2008 #50; Muñoz-Bernal, 2016 #51; Liu, 2016 #45}. The long history of consumption of organisms producing these close homologs of the 17 *Pichia sp.* proteins with no reports of allergenicity or toxicity offers strong general evidence for their safety in food.

Seven of the 17 *P. pastoris* proteins did not have significant alignments with proteins in the AOL database or among the known allergens or toxins in NCBI, while ten of the 17 had matches that are discussed below (Table 2).

The 1400 AA *P. pastoris* alpha aminoadipate reductase (Protein #1, Table S6) had no significant matches to proteins in AOL or to known allergens or toxins in NCBI.

The 768 AA *P. pastoris* cobalamin-independent methionine synthase (Protein #2a, Table S7) had a significant match (77% in the 80-mer search and one identical 8 AA match) to a protein in AOL: pollen allergen Sal k 3, a homologous protein from Russian thistle (*Salsola kali*, Table 2).

However, a phylogenetic tree (www.ebi.ac.uk/Tools/msa/clustalo/) constructed with sequences of *Pichia* host protein #2a, a homologous methionine synthase from *Saccharomyces cerevisiae*, Sal k 3

and homologues of Sal k 3 in commonly consumed plants. As expected, the phylogenetic tree in Figure 2a shows that allergen Sal k 3 has a much closer relationship to homologous proteins from plants to which it has the highest sequence identities. There is no evidence that these homologous proteins (89-91% identity) from plants are allergens or share IgE cross-reactivity with Sal k 3, thus it is unlikely that the match of Sal k 3 with the far more distantly related *P. pastoris* protein represents an allergy risk. In addition, the protein from the widely consumed *S. cerevisiae* has high sequence identity (77%) to the host protein #2a and is similarly lower in identity to homologues of Sal k 3 in edible plants.

The 780 AA *P. pastoris* aconitase (Protein #2b, Table S8) had low identity matches (not statistically significant) to homologous allergenic enzymes of *Aspergillus* sp. and the scabies mite. The low degree of identity suggests the risk of cross-reactivity is low, and the higher degree of similarity of the *P. pastoris* protein with aconitase from widely consumed *Saccharomyces* species, with no reports of aconitase allergies, suggests that the low identity matches to *Aspergillus* and scabies mite enzymes represent very low risk of allergenicity. Protein #2b had modest identity match to some proteins from toxic organisms, but not to any known toxins themselves. Higher scoring matches to proteins from widely consumed non-toxic yeasts and molds suggests that it is unlikely to pose any risk to humans.

The 679 *P. pastoris* AA transketolase (Protein #3, table S9) had no significant alignments to allergens or toxins in either AllergenOnline.org or the NCBI.

The 621 AA *P. pastoris* glycerol kinase (Protein #4, Table S10) had no significant alignments to any allergens, and only modest identity scores (<42%) with homologous proteins found in bacterial species that have evidence of toxicity to insects or vertebrates. The matched proteins however, are found in virtually all species and are not known to be toxins to animals (e.g. glycerol kinases).

The 510 AA *P. pastoris* catalase A (Protein #5a, Table S11) matched the homologous mold allergen Pen c 30 (*Penicillium citrinum*) in the 80-mer AA match with 60% identity to the best 80-mer

match (Figure 2b). However, the homologous protein from the closely related *Penicillium camemberti* is widely consumed in brie and other soft-ripened cheeses. It is 83% identical to Pen c 30 over its entire length {Galli, 2016 #30}. While there are two publications reporting allergenicity to *Penicillium camemberti* {Ovet, 2012 #56;Merget, 2008 #48}, there are no reports of food allergy associated with the enzyme. A protein from *Aspergillus oryzae*, which is widely used in Asia as koji in soy sauce and soy paste fermentation {Kitamoto, 2015 #40} also has a closer identity match to Pen c 30 than that represented by the *P. pastoris* catalase A. These widely consumed soybean-based foods do not have a history of allergenicity. Instead, they have catalases that have higher identity to Pen c 30 than it is to the *Pichia* protein, suggesting that the *Pichia* protein poses little risk of allergenicity. Finally, *Pichia* catalase is more closely related to catalase from the widely consumed *Saccharomyces* species discussed above (66% identity), with no known history of allergies. Collectively these results suggest that the risk of allergenic cross-reactivity of *Pichia* catalase to Pen c 30 is very low.

The 504 AA *Pichia* glucose 6 phosphate dehydrogenase (G6PD, Protein #5b, Table S12) shared 37% identity with the Bla g 3 protein from *Blattella germanica* (German cockroach). However the G6PDs from all widely consumed fungi (including baker's yeast, *Saccharomyces*, as well as mushrooms and morels) are far more similar to the *Pichia* protein than is Bla g 3. All have sequence similarities to Bla g 3 comparable to that of *Pichia* G6PD, indicating that there is a low risk of allergenic cross-reactivity between the *Pichia* protein and Bla g 3.

The 525 AA *Pichia* aldehyde dehydrogenase (ALDH, Protein #5c, Table S13) matched ALDH Cla h 10 (58% identity and $1e-126$ *E* score), and ALDH Alt a 10 (58% identity and $1e-124$ *E* score), which are identified as putative mold inhalation allergens based only on low-level IgE binding {Achatz, 1995 #4}. The sliding 80-mer window and a number of 8 AA identity matches were also significant for these proteins. An additional allergen, Lep d 13, from the storage mite, also yielded significant alignment using the 80-mer search window, although with a far lower score. ALDH are found in most organisms. A search of the NCBI Protein database without filters using the keyword

“allergen” identified numerous proteins with equal or greater similarity to Cla h 10 and Alt a 10, from food organisms commonly consumed as food by humans, including proteins from the commonly consumed yeast species, *Saccharomyces sp.* Moreover, the *Pichia* ALDH matched the ALDH from widely consumed *Saccharomyces* and other yeasts, with no history of ALDH allergies, at much higher similarity scores. We therefore conclude that the risk of cross reactivity between *Pichia* ALDH and either the mold or mite allergens is very low.

The 501 AA *Pichia* mitochondrial ALDH (Protein #6, Table S14) was also identified as having significant matches to the Cla h 10 and Alt a 10 proteins with 76% identities in optimal 80 AA alignments, as well as a number of 8 AA identity matches and overall FASTA alignments (Table S14). As noted above, the greater similarity of the matched putative allergenic proteins to commonly consumed proteins with no evident allergenicity demonstrates that these alignments do not suggest a significant risk of food allergy.

The 341 AA *P. pastoris* delta-aminolevulinate dehydratase (Protein #7a, Table S15) did not show significant alignments with any known allergens or toxins.

The 350 AA *P. pastoris* mitochondrial alcohol dehydrogenase (Protein #7b, Table S16) aligned with the homologous Can a 1 from the yeast *Candida albicans* by overall FASTA and 80-AA alignment, as well as 8-mer identity match. These proteins are ubiquitous mitochondrial alcohol dehydrogenases, as reflected by the results of a BLASTP search of the NCBI protein database, with similar sequence matches to numerous widely consumed food organisms without known allergenicity cross-reactivity.

The 342 AA *P. pastoris* malate dehydrogenase (Protein #7c, Table S17) aligned to a mold allergen Mala f 4 with 45% overall identity and a highest 80-mer identity of 70%, including 25 identical matches to 8 AA segments, indicating clear homology. Both Cand a 1.0101 and Mala f 4.0101 were partially purified and identified only with light IgE binding, leaving doubt as to their relevance for allergenicity {Shen, 1989 #63; Shen, 1991 #64; Onishi, 1999 #55}. Protein #7c also shared a minor identity match to the pea allergen Pis s 2, with highest score of 36.2% in a segment

of seven overlapping 80-AA alignments. A literature search did not reveal cross-reactivity between Mala f 4 and any *Pichia pastoris* proteins. A BLASTP search of the NCBI database without any keyword filter uncovered higher sequence identities to homologous proteins from a wide variety of sources, including the widely consumed *Saccharomyces* species, suggesting that the matches to Mala f4 and Pis s 2 do not represent a significant risk of allergenic cross-reactivity.

The 328 AA *P. pastoris* aldo/keto reductase (Protein #7d, Table S18) aligned with 24.6% identity to juniper allergen Jun o 4, and with only modest sequence identity matches to a few proteins from organisms associated with toxicity. These low scoring alignments suggested a low risk of allergenicity or toxicity. In contrast, the NCBI BLASTP search without keyword limit yielded higher-scoring identity matches to commonly consumed food species. Thus, there is little evidence of risk of allergic cross-reactivity from this protein.

The 248 AA triose phosphate isomerase (Protein #8, Table S19) had four relatively high scoring identity matches (50-53%) over near full-length alignments to homologous proteins from diverse sources (wheat, house dust mite, and shrimp). Matches were found using full-length FASTA, 80-mer alignment and 8 AA identity matches with AOL v16. Triose phosphate isomerase is a highly conserved enzyme; a search of the NCBI database using BLASTP without keyword filters yielded higher scoring identity matches to homologous proteins in species that are widely consumed without reported incidences of allergy, including a 71%-identity alignment to the enzyme from *Saccharomyces*. Thus, it is unlikely that the *Pichia* triose phosphate isomerase is an allergen.

The 161 AA *P. pastoris* cyclophilin (Protein #9a, Table S20) had significant alignments to a number of cyclophilins that have been identified targets of IgE from allergic subjects. We found relatively high identity matches, by overall FASTA, sliding 80-mer and 8-AA matches, to putative allergenic proteins from diverse sources, including Asp f 27, Der f 29, Mala s 6, Bet v 7, Asp f 11 and cyclophilins without IUIS-approved names from *Catharanthus sp.*, and carrot. Some cyclophilins were aligned using a BLASTP search of the NCBI database, filtered for “allergen”, “toxin” or “toxic” identified additional cyclophilins, but the same search, without those filters found better alignments

(>70% identity) to cyclophilins from many fungi including a *Saccharomyces* cyclophilin with 74% identity and a very significant *E* score (9e-88). In view of the closer similarity of the *P. pastoris* cyclophilin to its *Saccharomyces* homolog than to any known allergens, the matches to allergens are not likely to be indicative of an allergy risk for the *P. pastoris* cyclophilin.

The 154 AA *P. pastoris* cytosolic superoxide dismutase (Protein #9b, Table S21) aligned with significant scores to the 23 superoxide dismutase isoforms that have been identified from olive pollen as Ole e 5, which was identified as a putative allergen based on IgE binding using sera from a large number of olive-pollen allergic subjects. A cyclophilin from the widely consumed *Saccharomyces*, was a far better match to Protein #9b, with 79% identity over the entire length of 154 AA. No evidence was found for allergenic cross-reactivity between molds and olive pollen.

The 84 AA *P. pastoris* protein (Protein #10, Table S22) did not show significant alignments with any known or suspected allergens or toxins.

Bioinformatics searches for matches to allergens are intended to identify proteins that are already known to be allergens, or proteins that are nearly identical to an allergen from a different source to predict possible cross-reactivity. As noted by Aalberse {Aalberse, 2000 #2}, overall sequence identity alignments of less than 50% are unlikely to represent IgE cross-reactivity. The conservative limits of >35% identity in short segments of 80 amino acids are used to identify protein matches that might have any possibility of IgE cross-reactivity. The 8 AA identity matches have been largely discounted, but are still required by a number of regulatory agencies. If any match is identified to a known allergenic protein that aligns with higher identity than those limits, there is some expectation that serum IgE tests should be performed to identify IgE cross-reactivity {Goodman, 2008 #32}. However, IgE binding can be relatively low affinity or often may be restricted to a single epitope, which generally would not trigger a biological response and testing IgE binding in serum samples from a large number of allergic subjects is impractical. As a negative control in the bioinformatics analysis, we compared the proteins of interest to all known sequences in NCBI protein database using BLASTP. In each case, the negative-control analysis identified proteins

expected to be present in commonly consumed human foods, with no known allergenicity, that had far higher sequence-identity to the LegHb and *Pichia* proteins of interest than any of the matches we found between these proteins and known allergens.

Most of the putative allergens that had significant sequence similarity to the LegHb of the associated *Pichia* proteins were airway allergens from pollen, fungi or mites. None of those proteins nor any homologues of those proteins (e.g. cyclophilins, aldolase, alcohol dehydrogenases) have been reported to induce food allergy. Thus, the evidence from bioinformatics analyses, considered in its entirety, suggests little risk of allergenicity or toxicity from either the LegHb or the associated *Pichia* proteins.

Similarly, none of the results from analyses using BLASTP to search across the complete NCBI protein database suggested a significant risk of allergy or toxicity to consumers. Some of the *Pichia* proteins in the LegHb preparation were highly conserved intracellular enzymes with homologues that have been identified as putative allergens from a few organisms. But, in each case, homologues in organisms commonly consumed as human food had higher identity matches, without evidence of shared allergic reactivity.

Every significant sequence alignment between the 17 *Pichia* proteins of interest and proteins annotated with keywords “toxin” and “toxic” correspond to homologous proteins found in toxic organisms; none of these proteins were themselves known or suspected to be toxic. We could find no published evidence suggesting that any proteins with sequences similar to any of the 17 *Pichia* proteins might be toxic to humans or other mammals.

3.4 *In vitro* Pepsin digestibility results

To assess whether any of the proteins of interest might include digestion-resistant polypeptides, we carried out *in vitro* digestion assays at pH 2 in simulated gastric fluid, at pepsin-protein ratios of 10:1 (U/μg) or 1:1 (U/μg), comparing each protein of interest to three control proteins. Under these conditions two of the control proteins, hemoglobin (Figure 3a, 4a) and BSA

(Figure 3b, 4b), were digested rapidly, in 30 sec. The third control protein, OVA (Figure 3c, 4c) was considerably more stable, with more than 10% visually stainable full-length protein band remaining after 30 min digestion with 10 units of pepsin per microgram target protein (Figure 3c) and after 60 min at 1 unit per microgram target protein (Figure 4c). At both ratios, digestion of OVA left some residual polypeptides detectable by Coomassie stained SDS-PAGE as bands with apparent sizes of approximately 38 kDa after 60 min of digestion. The results of these control proteins were consistent with results from previously performed tests {Ofori-Anti, 2008 #54}, confirming the reliability of this digestion assay.

Two representative stained gels from digestion of LegHb at pH 2.0 with two ratios of pepsin:protein (Figure 3d, Figure 4d) demonstrated that LegHb was resistant to digestion in acid alone for 60 min (lane 3). However, LegHb was rapidly digested by pepsin in 2 min (lane 5) to below the 10% of the band intensity detected in a control digestion with quenched pepsin (lane 13). Additionally, all *Pichia* protein bands (Figure 1) were digested and no longer visible at time 0.5 min (lane 4 in Fig 3d and 4d), demonstrating they are rapidly digested as well. The results demonstrated that the LegHb protein and the associated *Pichia pastoris* proteins were more than 90% digested within 2 minutes in SGF plus pepsin at 37°C, at ratios of either 10 units or 1 unit of pepsin activity per 1 µg of total protein, based on Coomassie Blue staining detection. Thus, there is no added concern of risk based on pepsin-stability of this product preparation, according to CODEX (2003) guidelines for the allergenicity assessment {Codex, 2003 #16}.

4. Discussion

Bioinformatics comparison of the LegHb protein sequence compared with the protein sequences of known allergens did not identify any significant matches of sufficient concern to warrant serum IgE binding tests for possible cross-reactivity. The primary soybean allergens are major seed storage proteins, Gly m 5 (three isoforms of beta-conglycinin) and Gly m 6 (five isoforms of glycinin); and a number of other soybean proteins have been identified as potential allergens

(eight in the WHO/IUIS Allergen Nomenclature list; 15 in the OECD list of allergens, some of which have not had sequences determined). These allergens and potential allergens come from the edible bean, whereas leghemoglobin is expressed exclusively in the root nodule. Sequences of all published IgE binding proteins from soybean are included in AllergenOnline.org, which was the primary sequence search database used in this study. No significant sequence similarities were identified, indicating that LegHb is not homologous or significantly similar to any known allergen or IgE reactive protein in soybean. Though soybean is considered a commonly allergenic food source, soybean allergies are prevalent in infants and toddlers through IgE binding to proteins from the beans. Most of those allergic to soybeans become tolerant to soybean ingestion by the age of 10. Allergy to soy is uncommon in teenagers and adults {Savage, 2010 #61}. Clinical cross-reactivity among various foods from the legume family in children is rare {Bernhisel-Broadbent, 1989 #11}. Taken together, these results suggest that soy LegHb is unlikely to pose any risk of food allergy to soy-allergic or legume-allergic consumers.

Pichia pastoris is commonly used to produce recombinant proteins for cosmetics and pharmaceuticals. Bioinformatics searches with the 17 residual *Pichia* proteins found in Soy Leghemoglobin Preparation identified a few related protein sequences with sufficient similarity to exceed the CODEX suggestion for potential cross reactivity. However, the sequence-related putative allergens we identified in this search were not potent, common allergens, nor were any of them known to be allergenic when orally consumed. As the LegHb protein composes only up to 0.8% in the final plant-based meat products, the abundance of each *Pichia* protein is very trace amount. Moreover, comparison of the same *Pichia* proteins with all proteins in the NCBI Protein database found far more significant matches to proteins from commonly consumed fungi, including baker's yeast (*Saccharomyces* species). Impossible Foods' engineered *Pichia* production strain complies with the Organisation for Economic Co-operation and Development (OECD) criteria for Good Industrial Large Scale Practice (GILSP) microorganisms {OECD, 1992 #52}. It also meets the criteria for a safe production microorganism as described by Pariza and Foster {Pariza, 1983 #57}, Pariza and Johnson

{Pariza, 2001 #58}, and several expert groups {Berkowitz, 1989 #10;, 1992 #1;Council, 1990 #19;Jonas, 1996 #37;OECD, 1993 #53;FAO/WHO, 1996 #26}.

Although the overall CODEX paradigm for evaluating potential allergenicity has indicated taking a protein from a common source of allergy should trigger serum IgE testing using sera from those with allergy to the source; soybean (seeds) is a common allergenic food source, but root nodules are not. LegHb is not expressed in the bean. Thus there is no scientific rationale for tests using sera from soybean allergic subjects to determine whether there is IgE binding to LegHb. The other possibility is whether the protein is similar to any known allergen. The AA sequence of LegHb was evaluated by extensive bioinformatics comparison and there was no significant similarity to any allergenic protein. Additionally, the AA sequences of the residual *Pichia* proteins were evaluated by bioinformatics. There were no clear significant identity matches of *Pichia* proteins to known allergens. The LegHb and associated *Pichia* proteins were rapidly digested in simulated gastric fluid (pepsin). While these results do not prove that risks of allergy are non-existent, they demonstrate a very low likelihood of food allergy, and a lack of need for serum IgE tests or additional biological assays such as skin prick tests, basophil assays or challenges {Selgrade, 2009 #62;Codex, 2003 #16;Ladics, 2008 #42;Falcone, 2015 #25}.

A BLASTP search of the NCBI protein database for toxic proteins related by sequence to either LegHb or the associated *Pichia* proteins found no evidence suggesting that any of these proteins pose a risk of toxicity. While a number of organisms with known toxicity (e.g., *Bacillus sp.*, *Enterococcus faecalis*, *Streptomyces sp.*, *Clostridium sp.*) contained proteins with sequences similar to those of the *Pichia* proteins of interest, these proteins were ubiquitous and highly conserved across diverse species and are not themselves known or suspected to be toxic. As a further evaluation step, comparison of the sequence-related proteins from toxin-producing species to proteins from diverse non-toxic species identified far more closely related proteins from sources known to be safe and non-toxic.

In summary, our analysis found no evidence to suggest that food products containing the LegHb protein and associated minor components from *Pichia pastoris* pose any significant risk of allergy or toxicity to consumers.

Conflict of interest

RZF is an employee of Impossible Foods, the developer of the food product. REG has received consulting fees for work with Impossible Foods for recommendations for food safety evaluation procedures. The other authors declare that there are no conflicts of interest.

5. References

- [1] Everse, J., in: Lane, M. D. (Ed.), *Encyclopedia of Biological Chemistry*, Elsevier, New York 2004, pp. 354-361.
- [2] Abbaspour, N., Hurrell, R., Kelishadi, R., Review on iron and its importance for human health. *Journal of Research in Medical Sciences* 2014, 19.
- [3] Duff, S. M., Guy, P. A., Nie, X., Durnin, D. C., Hill, R. D., Haemoglobin expression in germinating barley. *Seed Science Research* 1998, 8, 431-436.
- [4] Appleby, C., Bergersen, F., Preparation and experimental use of leghaemoglobin. *Preparation and Experimental Use of Leghaemoglobin*. 1980, 315-335.
- [5] Fuchsman, W. H., Appleby, C. A., Separation and determination of the relative concentrations of the homogeneous components of soybean leghemoglobin by isoelectric focusing. *Biochimica et Biophysica Acta (BBA)-Protein Structure* 1979, 579, 314-324.
- [6] Hargrove, M. S., Barry, J. K., Brucker, E. A., Berry, M. B., Phillips, G. N. Jr., Olson, J. S., Arredondo-Peter, R., Dean, J. M., Klucas, R. V., Sarath, G., Characterization of recombinant soybean leghemoglobin a and apolar distal histidine mutants. *Journal of Molecular Biology* 1997, 266, 1032-1042.
- [7] Spohner, S. C., Müller, H., Quitmann, H., Czermak, P., Expression of enzymes for the usage in food and feed industry with *Pichia pastoris*. *Journal of Biotechnology* 2015, 202, 118-134.
- [8] Goodman, R. E., Vieths, S., Sampson, H. A., Hill, D., Ebisawa, M., Taylor, S. L., van Ree, R., Allergenicity assessment of genetically modified crops—what makes sense? *Nature Biotechnology* 2008, 26, 73-81.
- [9] Hammond, B., Kough, J., Herouet-Guicheney, C., Jez, J. M., Toxicological evaluation of proteins introduced into food crops. *Critical Reviews in Toxicology* 2013, 43, 25-42.
- [10] Ladics, G., Current codex guidelines for assessment of potential protein allergenicity. *Food and Chemical Toxicology* 2008, 46, S20-S23.
- [11] Codex, in: Commission, C. A. (Ed.), *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*. Codex Alinorm 03/34 2003.
- [12] Goodman, R. E., Performing IgE serum testing due to bioinformatics matches in the allergenicity assessment of GM crops. *Food and Chemical Toxicology* 2008, 46, S24-S34.
- [13] Mathesius, C., Barnett, J., Cressman, R., Ding, J., Carpenter, C., Ladics, G. S., Schmidt, J., Layton, R. J., Zhang, J. X. Q., Appenzeller, L. M., Safety assessment of a modified acetolactate synthase protein (GM-HRA) used as a selectable marker in genetically modified soybeans. *Regulatory Toxicology and Pharmacology* 2009, 55, 309-320.
- [14] Aalberse, R. C., Structural biology of allergens. *Journal of Allergy and Clinical Immunology* 2000, 106, 228-238.

- [15] Gieras, A., Cejka, P., Blatt, K., Focke-Tejkl, M., Linhart, B., Flicker, S., Stoecklinger, A., Marth, K., Drescher, A., Thalhamer, J., Mapping of conformational IgE epitopes with peptide-specific monoclonal antibodies reveals simultaneous binding of different IgE antibodies to a surface patch on the major birch pollen allergen, Bet v 1. *The Journal of Immunology* 2011, 186, 5333-5344.
- [16] Metcalfe, D. D., Astwood, J. D., Townsend, R., Sampson, H. A., Taylor, S. L., Fuchs, R. L., Assessment of the allergenic potential of foods derived from genetically engineered crop plants. *Critical Reviews in Food Science & Nutrition* 1996, 36, 165-186.
- [17] Fu, T.-J., Abbott, U. R., Hatzos, C., Digestibility of food allergens and nonallergenic proteins in simulated gastric fluid and simulated intestinal fluid A comparative study. *Journal of Agricultural and Food Chemistry* 2002, 50, 7154-7160.
- [18] Ofori-Anti, A., Ariyaratna, H., Chen, L., Lee, H. L., Pramod, S. N., Goodman, R. E., Establishing objective detection limits for the pepsin digestion assay used in the assessment of genetically modified foods. *Regulatory Toxicology and Pharmacology* 2008, 52, 94-103.
- [19] Bannon, G. A., Goodman, R. E., Leach, J. N., Rice, E., Fuchs, R. L., Astwood, J. D., Digestive stability in the context of assessing the potential allergenicity of food proteins. *Comments on Toxicology* 2002, 8, 271-285.
- [20] Goodman, R. E., Ebisawa, M., Ferreira, F., Sampson, H. A., van Ree, R., Vieths, S., Baumert, J. L., Bohle, B., Lalithambika, S., Wise, J., Taylor, S. L., AllergenOnline: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. *Molecular Nutrition & Food Research* 2016, 60, 1183-1198.
- [21] Siruguri, V., Bharatraj, D. K., Vankudavath, R. N., Mendu, V. V. R., Gupta, V., Goodman, R. E., Evaluation of Bar, Barnase, and Barstar recombinant proteins expressed in genetically engineered *Brassica juncea* (Indian mustard) for potential risks of food allergy using bioinformatics and literature searches. *Food and Chemical Toxicology* 2015, 83, 93-102.
- [22] Thomas, K., Aalbers, M., Bannon, G., 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-Guicheney, C., Holsapple, M., Ladics, G. S., Landry, T. D., MacIntosh, S. C., Rice, E. A., Privalle, L., Steiner, H. Y., Teshima, R., van Ree, R., Woolhiser, M., Zawodny, J., A multi-laboratory evaluation of a common in vitro pepsin digestion assay protocol used in assessing the safety of novel proteins. *Regulatory Toxicology and Pharmacology* 2004, 39, 87-98.
- [23] Collins, J. F., Fine, R., The enzymatic digestion of elastin at acidic pH. *Biochimica et Biophysica Acta (BBA)-Enzymology* 1981, 657, 295-303.
- [24] Crévieu-Gabriel, I., Gomez, J., Caffin, J.-P., Carré, B., Comparison of pig and chicken pepsins for protein hydrolysis. *Reproduction Nutrition Development* 1999, 39, 443-454.
- [25] Astwood, J. D., Leach, J. N., Fuchs, R. L., Stability of food allergens to digestion in vitro. *Nature Biotechnology* 1996, 14, 1269-1273.
- [26] FAO/WHO, Rome, Italy 2001.

- [27] Xie, Y., Xu, D., Cui, W., Shen, W., Mutation of *Arabidopsis* HY1 causes UV-C hypersensitivity by impairing carotenoid and flavonoid biosynthesis and the down-regulation of antioxidant defence. *Journal of Experimental Botany* 2012, 63, 3869-3883.
- [28] Kim, H. J., Sung, M.-K., Kim, J.-S., Anti-inflammatory effects of glyceollins derived from soybean by elicitation with *Aspergillus sojae*. *Inflammation Research* 2011, 60, 909-917.
- [29] Dumez, M.-E., Herman, J., Campisi, V., Bouaziz, A., Rosu, F., Luxen, A., Vandenberghe, I., De Pauw, E., Frere, J.-M., Matagne, A., The proline-rich motif of the proDer p 3 allergen propeptide is crucial for protease-protease interaction. *PloS one* 2013, 8, e68014.
- [30] Rodriguez-Perez, R., Caballero, M. L., Gonzalez-Munoz, M., Rodriguez-Mahillo, A., Moneo, I., Cloning and expression of a biologically active *Anisakis simplex* allergen Ani s 1 in the yeast *Pichia pastoris*. *Molecular and Biochemical Parasitology* 2007, 154, 115-118.
- [31] Marazuela, E., Rodriguez, R., Barber, D., Villalba, M., Batanero, E., Hypoallergenic mutants of Ole e 1, the major olive pollen allergen, as candidates for allergy vaccines. *Clinical & Experimental Allergy* 2007, 37, 251-260.
- [32] Yasuhara, T., TAKAI, T., Yuuki, T., Okudaira, H., Okumura, Y., Cloning and expression of cDNA encoding the complete prepro-form of an isoform of Der f 1, the major group 1 allergen from house dust mite *Dermatophagoides farinae*. *Bioscience, Biotechnology, and Biochemistry* 2001, 65, 563-569.
- [33] Kastenmüller, K., Espinosa, D. A., Trager, L., Stoyanov, C., Salazar, A. M., Pokalwar, S., Singh, S., Dutta, S., Ockenhouse, C. F., Zavala, F., Full-length *Plasmodium falciparum* circumsporozoite protein administered with long-chain poly (I·C) or the Toll-like receptor 4 agonist glucopyranosyl lipid adjuvant-stable emulsion elicits potent antibody and CD4+ T cell immunity and protection in mice. *Infection and Immunity* 2013, 81, 789-800.
- [34] Banerjee, H. n., Verma, M., Search for a novel killer toxin in yeast *Pichia pastoris*. *Plasmid* 2000, 43, 181-183.
- [35] Chessa, R., Landolfo, S., Ciani, M., Budroni, M., Zara, S., Ustun, M., Cakar, Z. P., Mannazzu, I., Biotechnological exploitation of *Tetrapisipora phaffii* killer toxin: heterologous production in *Komagataella phaffii* (*Pichia pastoris*). *Applied Microbiology and Biotechnology* 2017, 101, 2931-2942.
- [36] Comitini, F., Mannazzu, I., Ciani, M., *Tetrapisipora phaffii* killer toxin is a highly specific beta-glucanase that disrupts the integrity of the yeast cell wall. *Microbial cell factories* 2009, 8, 55.
- [37] Moyad, M. A., Brewer's/baker's yeast (*Saccharomyces cerevisiae*) and preventive medicine: Part II. *Urologic Nursing* 2008, 28, 73.
- [38] Muñoz-Bernal, E., Deery, M. J., Rodríguez, M. E., Cantoral, J. M., Howard, J., Feret, R., Natera, R., Lilley, K. S., Fernandez-Acero, F. J., Analysis of temperature-mediated changes in the wine yeast *Saccharomyces bayanus* var *uvarum*. An oenological study of how the protein content influences wine quality. *Proteomics* 2016, 16, 576-592.

- [39] Liu, J.-J., Kong, I. I., Zhang, G.-C., Jayakody, L. N., Kim, H., Xia, P.-F., Kwak, S., Sung, B. H., Sohn, J.-H., Walukiewicz, H. E., Metabolic engineering of probiotic *Saccharomyces boulardii*. *Applied and Environmental Microbiology* 2016, 82, 2280-2287.
- [40] Galli, B. D., Martin, J. G. P., da Silva, P. P. M., Porto, E., Spoto, M. H. F., Sensory quality of Camembert-type cheese: Relationship between starter cultures and ripening molds. *International Journal of Food Microbiology* 2016, 234, 71-75.
- [41] Ovet, H., Ergin, C., Kaleli, I., Investigation of mold fungi in air samples of elementary schools and evaluation of allergen-specific IgE levels in students' sera. *Mikrobiyoloji Bulteni* 2012, 46, 266-275.
- [42] Merget, R., Sander, I., Rozynek, P., Heinze, E., Imoehl, M., Raulf-Heimsoth, M., Bruening, T., Occupational Immunoglobulin E-Mediated Asthma due to *Penicillium camemberti* in a Dry-Sausage Packer. *Respiration* 2008, 76, 109-111.
- [43] Kitamoto, K., Cell biology of the Koji mold *Aspergillus oryzae*. *Bioscience, Biotechnology, and Biochemistry* 2015, 79, 863-869.
- [44] Achatz, G., Oberkofler, H., Lechenauer, E., Simon, B., Unger, A., Kandler, D., Ebner, C., Prillinger, H., Kraft, D., Breitenbach, M., Molecular cloning of major and minor allergens of *Alternaria alternata* and *Cladosporium herbarum*. *Molecular Immunology* 1995, 32, 213-227.
- [45] Shen, H.-D., Choo, K. B., Tang, R. B., Lee, C. F., Yeh, J.-Y., Han, S. H., Allergenic components of *Candida albicans* identified by immunoblot analysis. *Clinical & Experimental Allergy* 1989, 19, 191-196.
- [46] Shen, H.-D., Choo, K.-B., Lee, H.-H., Hsieh, J. C., Lin, W.-L., Lee, W.-R., Han, S.-H., The 40-kilodalton allergen of *Candida albicans* is an alcohol dehydrogenase: molecular cloning and immunological analysis using monoclonal antibodies. *Clinical & Experimental Allergy* 1991, 21, 675-681.
- [47] Onishi, Y., Kuroda, M., Yasueda, H., Saito, A., Sono-Koyama, E., Tunasawa, S., Hashida-Okado, T., Yagihara, T., Uchida, K., Yamaguchi, H., Two-dimensional electrophoresis of *Malassezia* allergens for atopic dermatitis and isolation of Mal f 4 homologs with mitochondrial malate dehydrogenase. *European Journal of Biochemistry* 1999, 261, 148-154.
- [48] Savage, J. H., Kaeding, A. J., Matsui, E. C., Wood, R. A., The natural history of soy allergy. *Journal of Allergy and Clinical Immunology* 2010, 125, 683-686.
- [49] Bernhisel-Broadbent, J., Taylor, S., Sampson, H. A., Cross-allergenicity in the legume botanical family in children with food hypersensitivity: II. Laboratory correlates. *Journal of Allergy and Clinical Immunology* 1989, 84, 701-709.
- [50] OECD, *Safety Considerations for Biotechnology* 1992, p. 45.
- [51] Pariza, M., Foster, E., Determining the safety of enzymes used in food processing. *Journal of Food Protection* 1983, 46, 453-468.
- [52] Pariza, M. W., Johnson, E. A., Evaluating the safety of microbial enzyme preparations used in food processing: update for a new century. *Regulatory Toxicology and Pharmacology* 2001, 33, 173-186.

[53] Berkowitz, D., Maryanski, J., *Codex Alimentarius Commission. Sess. 18, Geneva (Switzerland), 3-14 Jul 1989* 1989.

[54] *EU Scientific Committee for food* 1992, pp. 13-22.

[55] Council, I. F. B., Chapter 4: Safety evaluation of foods and food ingredients derived from microorganisms. *Regulatory Toxicology and Pharmacology* 1990, 12, S114-S128.

[56] Jonas, D. A., Antignac, E., Antoine, J.-M., Classen, H.-G., Huggett, A., Knudsen, I., Mahler, J., Ockhuizen, T., Smith, M., Teuber, M., The safety assessment of novel foods: guidelines prepared by ILSI Europe novel food task force. *Food and Chemical Toxicology* 1996, 34, 931-940.

[57] OECD, *OECD safety evaluation of foods derived by modern biotechnology: concepts and principles* 1993.

[58] FAO/WHO, *Joint FAO/WHO Expert Consultation on Biotechnology and Food Safety* 1996, p. 61.

[59] Selgrade, M. K., Bowman, C. C., Ladics, G. S., Privalle, L., Laessig, S. A., Safety assessment of biotechnology products for potential risk of food allergy: implications of new research. *Toxicological Sciences* 2009, 110, 31-39.

[60] Ladics, G., Van Bilsen, J., Brouwer, H., Vogel, L., Vieths, S., Knippels, L. M. J., Assessment of three human FcεRI-transfected RBL cell-lines for identifying IgE induced degranulation utilizing peanut-allergic patient sera and peanut protein extract. *Regulatory Toxicology and Pharmacology* 2008, 51, 288-294.

[61] Falcone, F. H., Alcocer, M. J., Okamoto-Uchida, Y., Nakamura, R., Use of humanized rat basophilic leukemia reporter cell lines as a diagnostic tool for detection of allergen-specific IgE in allergic patients: time for a reappraisal? *Current Allergy and Asthma Reports* 2015, 15, 67.

Figure captions

Figure 1. Coomassie Brilliant Blue Stained SDS-PAGE Gel showing a production lot of Leghemoglobin that includes the 17 identified residual *Pichia pastoris* proteins. A production batch of LegHb was applied to an SDS-PAGE gel using reducing conditions. The MW of the marker proteins (lane 1) are indicated as is the position of LegHb (lane 2) and the major bands from the *Pichia* host.

Figure 1.

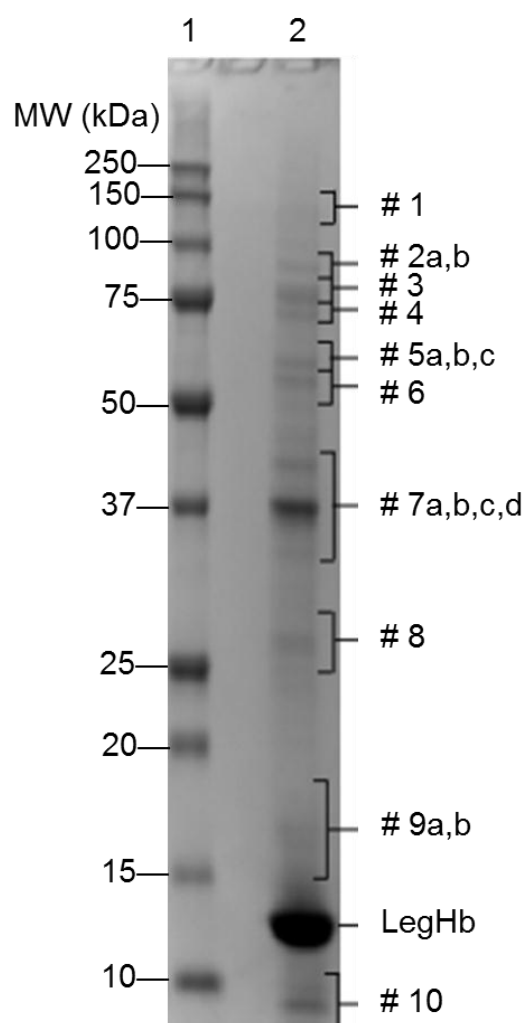


Figure 2. Two phylogenetic trees as examples comparing host *Pichia pastoris* proteins #2a (a) and #5 (b) to sequences of matched allergens from Allergenonline.org, and sequences of proteins from edible plants, and edible fungi *Saccharomyces sp.* and *Aspergillus oryzae*. The edible plant proteins and fungal proteins were identified by BLASTP of the NCBI non-redundant database and a close homologue of the *Pichia* protein from an edible fungal source.

Figure 2.

Figure 2a

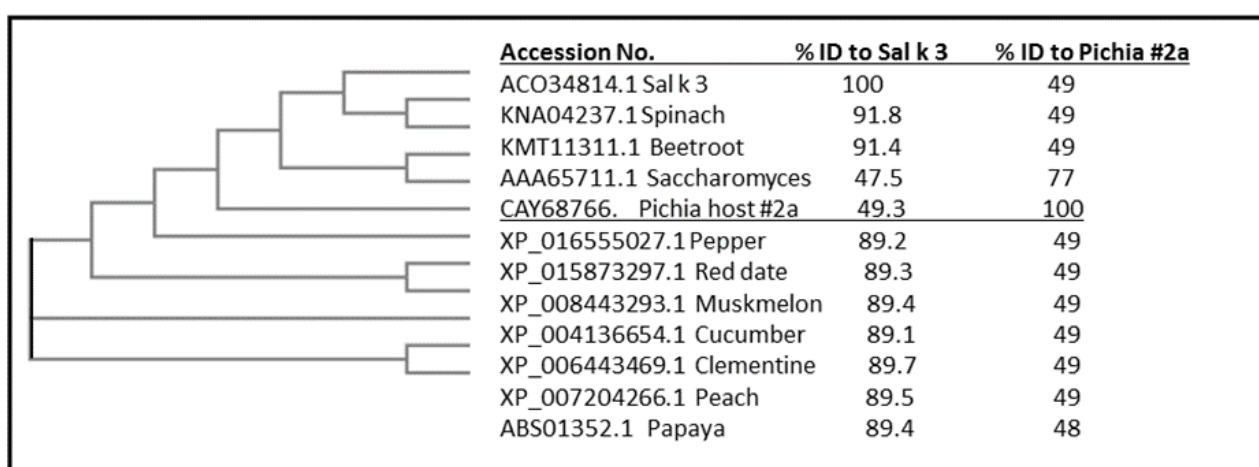


Figure 2b

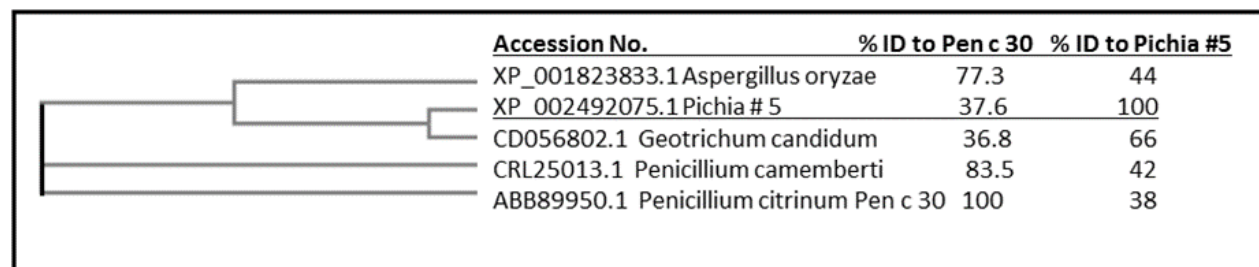


Figure 3. Coomassie Brilliant Blue Stained SDS-PAGE Gel Showing the Digestion of control samples, Bovine Hemoglobin (a), BSA (b) and OVA (c), and LegHb preparation (d) in SGF at the ratio of 10 Units pepsin per μg test protein (pH 2.0). All proteins were loaded 1.47 μg per lane as pre-digestion concentration. Lane M, molecular weight marker; Lane 1, protein control at 0 min; Lane 2, protein control at 60 min; Lane 3-10, digestion at 0, 0.5, 2, 5, 10, 20, 30, 60 min; Lane 11, pepsin control at 0 min; Lane 12, pepsin control at 60 min; Lane 13, 10% of undigested protein.

Figure 3.

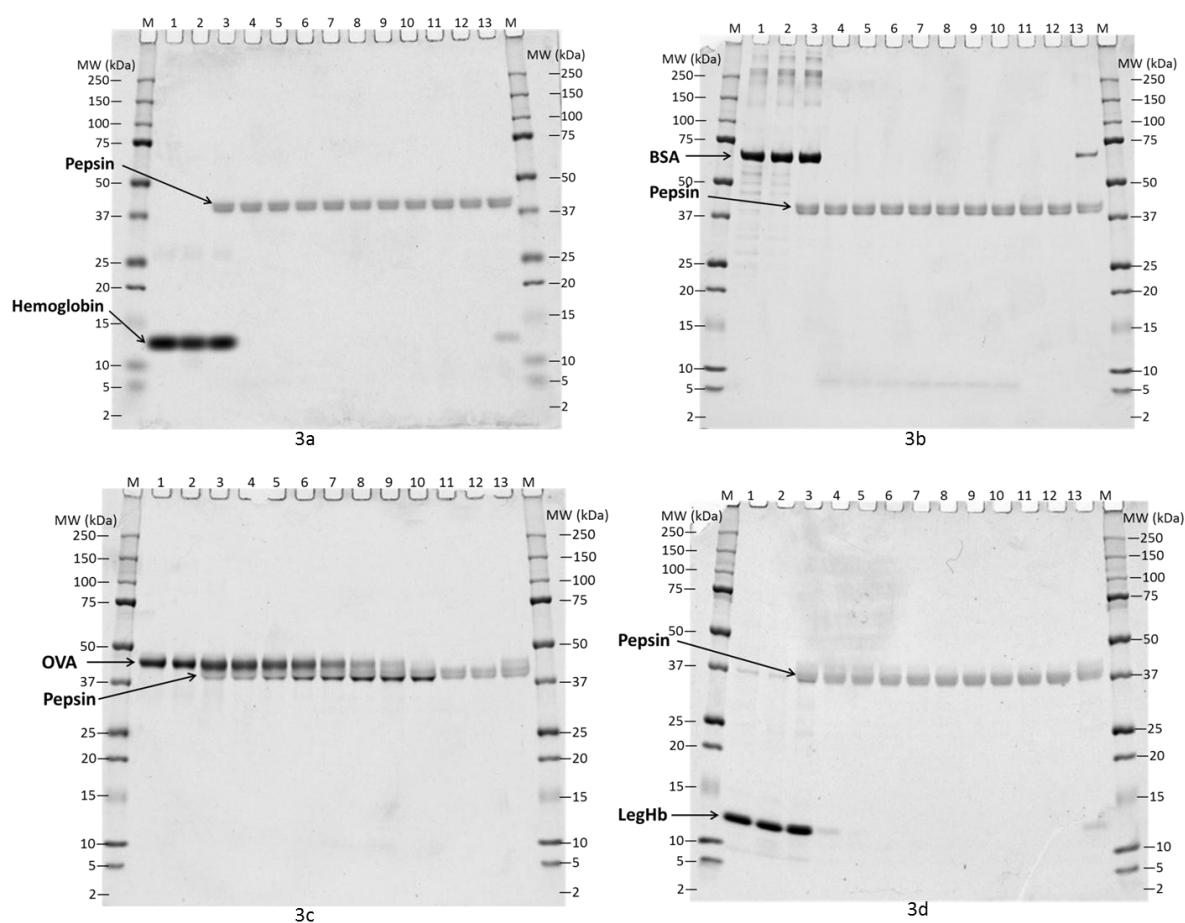


Figure 4. Coomassie Brilliant Blue Stained SDS-PAGE Gel Showing the Digestion of control samples, Bovine Hemoglobin (a), BSA (b) and OVA (c), and LegHb preparation (d) in SGF at the ratio of 1 Unit pepsin per μg test protein (pH 2.0). All proteins were loaded 1.47 μg per lane as pre-digestion concentration. Lane M, molecular weight marker; Lane 1, protein control at 0 min; Lane 2, protein control at 60 min; Lane 3-10, digestion at 0, 0.5, 2, 5, 10, 20, 30, 60 min; Lane 11, pepsin control at 0 min; Lane 12, pepsin control at 60 min; Lane 13, 10% of undigested protein.

Figure 4

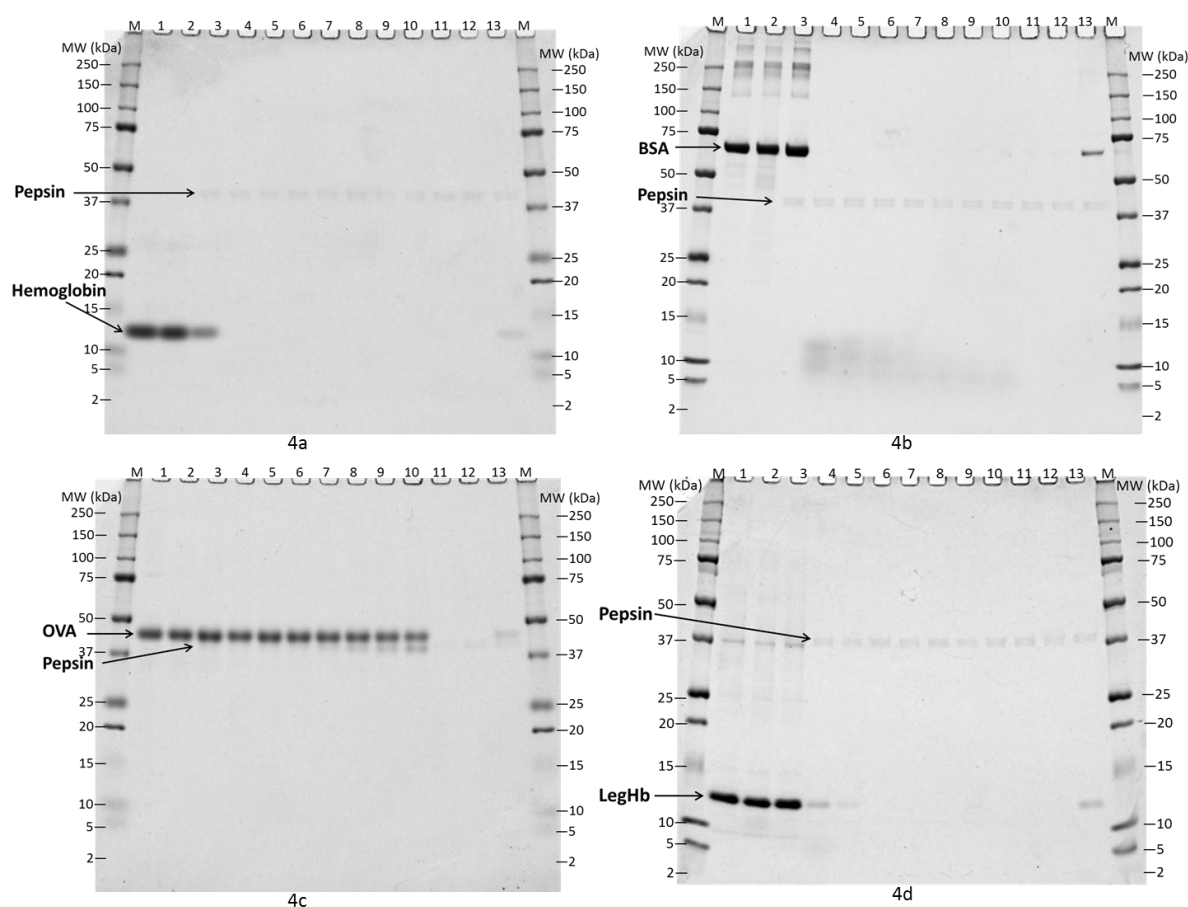


Table 1 Amino acid (AA) sequence of the soybean leghemoglobin protein produced in *Pichia pastoris*.

Organism	Hemoglobin class GI#: Accession#	Native protein sequence
<i>Glycine max</i>	Leghemoglobin c2 (LegHb) GI:126241 Acc: P02236.2	MGAFTEKQEALVSSSFEAFKANIPQYSVVFYTSILEKAPAAKDLFSFLSNGVDPSNPGLTGH AEKLFGLVRDSAGQLKANGTVVADAALGSIHAQKAITDPQFVVVKEALLKTIKEAVGDKW SDELSSAWEVAYDELAIAIKKAF

Table 2 Summary of sequence alignments for LegHb and the 17 host proteins identified in AllergenOnline searches.

Query Protein			AOL Results							
			Full FASTA			80-mer			8-mer	
			No. of matches			Matched Allergen ⁴	No. of hits ID > 35%	Best ID ⁵	Matched allergen	No. of hits
			E score		ID ³ >50%					
			<10	<1						
No. ¹	protein name (GI#) [Accession]	No. of AA ²								
n/a	LegHb (126241) [P02236.2]	145	20+	12	0	n/a ⁵	0	n/a	n/a	0
1	Alpha aminoadipate reductase (238030060) [CAY67983.1]	1400	4	0	0	n/a	0	n/a	n/a	0
2a	Cobalamin-independent methionine Synthase (238030843) [CAY68766.1]	768	12	1	0	Sal k 3	630	77.5%	Sal k 3	21
2b	Aconitase (254564667) [XP_002489444.1]	780	14	2	0	n/a	0	n/a	n/a	0
3	Transketolase (238030057) [CAY67980.1]	679	20+	0	0	n/a	0	n/a	n/a	0
4	Glycerol kinase (238034027) [CAY72049.1]	621	20+	0	0	n/a	0	n/a	n/a	0
5a	Catalase A (254569930) [XP_002492075.1]	510	20+	4	0	Pen c 30	214	60%	n/a	0
5b	G6PD (238031000) [CAY68923.1]	504	6	1	0	Bla g 3	10	37%	n/a	0
5c	Hypothetical protein PAS (238031215) [CAY69138.1]	525	20+	8	2 (Cla h 10, Alt a 10)	Cla h 10	442	72.5%	Cla h 10	41
						Alt a 10	442	72.5%	Alt a 10	27
						Lep d 13	1	35.4%		
6	Mitochondrial ALDH (238033249) [CAY71271.1]	501	20+	8	2 (Cla h 10, Alt a 10)	Cla h 10	422	76.2%	Cla h 10	41
						Alt a 10	422	76.2%	Alt a 10	25

Received: 17-04-2017; Revised: 09-08-2017; Accepted: 30-08-2017

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/mnfr.201700297](#). This article is protected by copyright. All rights reserved.

7a	Delta-aminolevulinate dehydratase (238033645) [CAY71667.1]	341	9	0	0	n/a	0	n/a	n/a	0
7b	Mitochondrial alcohol dehydrogenase isozyme III (238031179) [CAY69102.1]	350	12	1	1 (Cand a 1)	Cand a 1	271	85%	Cand a 1	49
7c	Malate dehydrogenase (238034064) [CAY72086.1]	342	18	4	1 (Mala f 4)	Mala f 4	216	70%	Yeast Mala f 4	25
						Pis s 2	7	36.2%		
7d	Putative protein, unknown function (238033788) [CAY71810.1]	328	10	1	0	n/a	0	n/a	n/a	0
8	Triose phosphate isomerase (238032989) [CAY71012.1]	248	10	5	4 (Tri a 31, two isoforms of Der f 25, Cra c 8)	Tri a 31	169	62.5%	Der f 25	17
						Der f 25.0101 (isoform)	169	60.0%	Der f 25	17
						Der f 25.0201 (isoform)	169	60.0%	Cra c 8	11
						Cra c 8	169	57.5%	Tri a 31	17
9a	Hypothetical protein (cyclophilin) PP7435 (328350030) [CCA36430.1]	161	17	7	7 (PPlase ⁶ from fungus: Asp f 27, Der f 29, Mala s 6, Cat r 1, PPlase from birch: Bet v 7, from carrot <i>Daucus carota</i> , and from Fungus: Asp f 11)	Mala s 6	82	87.5%	Asp f 27	23
						Asp f 27	82	85%	Der f 29	23
						Cat r 1	82	81.3%	Bet v 7	11
						Der f 29	82	80%	Cat r 1	19
						Asp f 11	82	80%	(Unassigned by IUIS) PPlase from <i>Daucus carota</i>	10
						Bet v 7	82	80%	Mala s 6	31
						(Unassigned by IUIS) PPlase from <i>Daucus carota</i>	82	78%	Asp f 11	17
9b	Cytosolic superoxide dismutase (238034030) [CAY72052.1]	154	20+	23	23 (isoforms of Ole e 5)	Ole e 5 23 isoforms	75	60% - 55%	n/a	0
10	Mitochondria ATPase Inhibitor (238029769) [CAY67692.1]	84	17	4	0	n/a	0	n/a	n/a	0

¹Protein band numbers correspond to the numbers labeled on the stained gel in Figure 1;

²AAs, amino acids;

³ID, identity (%);

⁴Allergen name in IUIS allergen list unless denoted as unassigned by IUIS;

⁵n/a: not available or no answer.

⁶PPIase: peptidylprolyl isomerase.

Table S1 Amino acid sequences of 17 proteins from *Pichia pastoris* (synon. *Komagataella phaffii*) in the product. Protein band numbers are designated as in Figure 1. Both GI and Accession numbers are listed as NCBI modified their format during this study and they no longer list GI numbers.

Organism	Protein, GI, band# and % coverage by LC-MS/MS	Native protein sequence
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	Band #1: Alpha amino adipate reductase (1400 AA) GI:238030060 Acc: CAY67983.1 32%	MSEENLNYWANILDGPTLSVLPRDYNRPVAGKVEANKTFDISDILPFLNKANEPVSTQFTAPLAVFAVLVYRLTGDDDDIVILT DSPKKQNLFPVRLQVDPSPKSFVDVSKQVGEQYLESLEERATPLKDIVTHLKEKSKLPNYPPIRFLSFQAKKVQQLSTLVEGST RDLAIFLENNTSINIYVNSLLYTHNRIAYFSQQFSSFIDEVNAKAPETPIGKISLLTEQQSKLLPDPTANLWDSGYRGAIQDFSDN AEKFPDRTCVVETKSLFNPNSQTRTFTYKQIDQASNIVGNVYLVTGKIRGDDVMMIAYRQVGLMVAVMGVKAGATFSVID PAYPPARQNVYLQVAKPAGLIVLEKAGVLDQLVEDYIKNELSLVSRISNLKIEADGNVLGGDDVGDKDALYDYQKFTRTRTGV VGPDSNPTLSFTSGSEGIPKGVLRHFSLAYYFPWMSKTFNLSENDFKFTMLSGIAHDPRIQRDMFTPLFLGAQLLIPTSDDIGT PGKLAEWMMQTYGATVTHLTPAMGQLLSAQATKEIPSLHHAFFVGDILTKRDLRLQIAQNVNINMYGTTETQRAVSYPEI PSRAQDSTFLEVQKDIMPAGKGMHNQVLLVNRHDSKTKAIGEVGEIYVRAAGLAEQYRGQPDNLKEKFPVNPWFVSPSK WVEEDKKISKDEPWREFYLGPRDRLYRTGDLGRYLTGDCVSGRADDQVKIRGFRIELGEIDTHISRHPILIRQNVTLVRRDK DEEPILISYVVPKTEPELENFKSSDDLDLNDPIVKSLLYRELKDLKAHLKKTASYAIPITVMAKLPNPNKGVDKPKLPFP DTVQLAAVAQKSSAEVDDSEFTTTELQIKDLWLQVLPNPPASISLEDSFFDLGGHSLATRMIFELRRKLAVDLPLGTIFKHPTV KLFAAEVDRVKNQDEVDQADNKQESTSAGSDEQVVDYFQDAKDLVSSQLLSDYSKRLASNAELINFLTGTATGFLGSYILKDL LERDLVQVYAHVRAKDEESGLERLNTGKVIWNEEWSRIKVIADLSKDKLGLSGEKYAEALNTIDLIHNGALVHWVY PYSKLRDANVISTINVLNLAASGKPKQGFVSSTSLDTEHYITLSDTLTEQGEDGIPESDDLLGSSKGLGTGYGQSKWAAEYII RRAFERGLRGAIIRPYVTGHSRTGACNTDDFLLRMLKGAELGKLPNISNTVMNVPVDHVALVVTASSLHPTAEEGHCVV QVTGHPRIRFNEFLNALNDYGEVVKLTDYVEWKRDLERFVVDQSKDSALYPLLHFVLDNLPQDTKAPELDDKNKADILSGDT RWTGYDGSKGRGVDASQGTGIYAIYLIKTFGLPPPSKEGKKPLPEIIESESLKLIKEGAGARTSAA
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#2a: Cobalamin-independent methionine synthase (768 AA) GI:238030843 Acc: CAY68766.1 77%	MVQSSVLGFPRIAGFRELKKTTEAYWSGKVGKDELKFKVKEIRENNWKLQKAAGVDVIANSDFSYYDQVLDLSLNAIPER YTKYELDPIDTLFAMGRGLQRKATDSEKAVDVTALEMVKWFDSNYHYVRPTFSHSTEFKLNQKQPVDEYLEAKKLGIERTPV VVGPPSYFLFGKADKDSLDEPISLLEKILPVYAEALLAKLSAAGATSVQIDEPILVLDLPEKVQAAFKTAYEYLANAKNIPKLVVAS YFGDVRPNLASIKGLPVHGFHDFVRAPEQFDEVVAALTAEQVLSVGIIDGRNIWKADFSEAVAFVEKAIAALGKDRVIVATS SSLHTPTVDLTNEKDLSEIKNWFSAFATQKLDEVVVAKAVSGEDVKEALSNAAIKSRKDSAITNDADVQKVVQDSINEKLSS RAAFAFERLAQAQKGFNLPPLFTTTTIGSFQTKDIRINRNKFTKGEITAEQYDTFIKSEIKVVRFEQIEIGDLVLVHGEPRNDM VQYFGEQLKGFAFTTNGWVQSYGSRVYRPPVVGDSVRPHAMSVKESVYQAISITKKPMKGMILTGPITVLRWVSFPRNDVS QKVQALQLGLALRDEVNDLEAASVEVIQVDEPAIREGLPLRSQERSDYLYAAESFRIATSGVKNITQIHSHFCYSOLDPNHI KALDADVVSIEFSKDDPNYIQESFNYPNHIHGLGLFDIHSRIPRSKEEFARIAGEILKVPYADKFWVNPDCGLKTRGWEEVVRASL TNMVEAAKTYREKYAQN
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#2b: Aconitase (780 AA) GI:254564667 Acc: XP_002489444.1 57%	MLSARRVLAKINSRGLATVSLTKDSLVMENLLEKGNINYKQLDNNVIVKERLGRPLTYAEKLYGHLDKPEHQDIERGVS YKLPRDIACQDATAQMAILQFMSAGMPSVATPTTVHCDHLIAQKGGGAADLERAIRLNKEVYDFLATACAKYINIGFWKP GSGIHHQIVLENYAFPGELLIGTDSHTPNAGGLGQLAIGVGGADAVDVMAGLPWELKAPKIIIGVGLTGRMNGWTSPPKDIILK LAGITTVKGGTGAIVEYFGDGVDTFSCGMATICNMGAIEGATTSVFPFNNSMVDFLDATGRSEIGEFKAVFQKEYLSADPG CEYDQVIEIDLNTLEPHINGPFTPDLATPVSKMKKEVAVANDWPLEVKVGLIGSCTNSSEYDMTRAASIEDAASHGVKAKSLT VTPGSEQIRIARDGQLKFTDFGGSVLNACGPGCIGQWDRQDIKKGDKNTIVSSFNRFSTRNDGNPATHAFVASEPMV TAYAIAGDLRFNPLTKDLKDKGNEFLKDPVGVGLPVGRYDPGENTYQAPPEDRASVEVVISPSDDLRLTLPFPQWPDGK DAERLPILIKSVGKTTTDHISMAGPWLYKRGHLQNISNNYMGAINAENGANNVKNHYTVGSYGVPTDAAAYRDNQVVKW VVGIGENFEGSSREHAALPRYLGGFAITKFSARIHETNLKQGLLPNFTDPAAYDRIQPDDEVDILGTELAPGNKVTLRV HPADGSPWTETPLSHTYNAEQIEWFKYGSALNNMAAVKASK
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#3: Transketolase (679 AA) GI:238030057 Acc: CAY67980.1 63%	MSDLLAINTIRLLAVDTVAKANSNGHPGAPMGLAPAAHTLFKQMRFNPRNPAWINRDRFLVSNHGACALLYTMLFLGYDY TIDDLKSRQLNSKTPGHPEALPGVEVTTGPLQGIANAVGLAIAQAQLAATYKNRPFVLSNYAFGLDGCLQEGVAQE AISLAGHLGDGLKLIADFDDNQISIDGDTNVSFTEVPAKFRAQGWELVSVKDGNDLLEGIAAALAKAKTTNKPTLIRLPTII GYGSLQQGTGHVGHGSPKPDIDIKKKKFGDFPEQNFVVPREVTSYAKHVADNQQVEVEWKNLLTAYTKEYPELQELH RRLDGKLPEWQKALPTYVDDKPVASRKLSEIVLTSIEKELPELVGGSADLTGSNLTRWPDVDFQPKSTGLGDFSGRYFRF GVREHGMGAIIINGISAYGANFKAYGGTFLNFVSYASGAVRLSALSHPHIIWVATHDSIGLGDGDPTHQPIETLAHLRALPNL MVWRPADGNETSAAYLRAIESKHPTSIALLTRQNLPLEHSSIEKAAKGGYTVYPVENPDILVASGSEVSIADGAKKLTGEGV KASVVSIPDFFTFDSQARSYQLSVLPDGVPMISVEVMSTFGWSKYSHEQGINRFGISGPGPEIYKFEFFTAEGVADRASKVV QFYKGKELSLPLNKAFESVHA
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#4: Glycerol kinase (621 AA) GI:238034027 Acc: CAY72049.1 61%	MGKDYTPLVATIDIGTTSTRAILFDYHQGEVAKHQEYSTSAQDDIKRKRQIISSEGISLTVSDDLVEVESVDNKAAPTQFPQP GWVECRPSHILANAVQCLAACLTVMENKNLDRDEKNKYKISIGVANMRETTVVWSKKTGKPLYNGIVWNDTRNNDIVDE YTAKEYSEKEREEMRLTLCGPISYFSAFKFRLKLVPEVKQAYDNADGDMFGTIDSWLIYHLTNEKSHVTDVNTASRTNF MNIETNKYDDRLKFWVDVDTSKVILPEIRSSAEVYGHFKVPHLESIGYVESYLTDDALLLETIEGAPLAGCLGDQASALVGLQ AVRKGDAKCTYGTGAFLLYNTGDQTLSEHGALTTVGYWFPGLDESEDGKHSSKPYALEGSIYAVGSSVQWLRDLNRLISK AQDVGPLASQVDNSGGVVFPVAFSGFLFAPYWDNSRGTIFGLTQYTSASHIARAALLEGVCFTQTRAILKAMISDAGASADFL ESSKATGHNPVSLVLDGGMKSDDEMMLQIADILGPCVTVRRSINPECTALGAAIAAGFVPKEDRIWGSLECTEAILGN KMYLAAGNTSLDFKATLSDEVRRKEWRLWENAIKAKGWLKDTA
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#5a: Catalase A (510 AA) GI:254569930 Acc: XP_002492075.1 86%	MSQPPKWTTSSNGAPVSDVATERATFDNANHANNAKPVGPLLQDFQLIDSLAHFDRERIPERVVHAKGAGAFGEFEVTD DISDVCAAKFLDTIGKKTIRFTSTVGGEGKSADSARDPRGFSTKFYTEGNLDLVYNNPTPIFIRDPSKPFPHIHTQKRNPAT NLKADANMFWDYLVNNGQESIHQVMYLFSDRGTPASLRKMNNGSYGHTYKWNKKGEWVYVQHFKSDLVGVVNFNNEAG KLAGEDPDYHTGDLFNAIERGEYPSWTCYIQTMTQEQAAKQPFVFDLTKVWPHKDFPLRRFGKFTLNENPNKNYFAVEVQA AFSPSTHIPSMPQADPVLQSRFLSYDPDTHRRLGVNYQQIPVNCVPVFTPQMRDGSMTVNGNLGTSPPNYKSSFCPFST EAQIQTNSTHTEPEVLAAHTEKFWHGGILDSKSYDFEQPRALWVKVFGKTPGQQRNFCNPNVAVHVAAANHEIQDRVFEYFSK VYPEIGDQIRKEVLQSLPRGDSAAAL

<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#5b: Glucose 6 phosphate dehydrogenase (G6PD) (504 AA) GI:238031000 Acc: CAY68923.1 62%	MTDTKAVEFVGHTAIVVFGASGDLAKKKTFFALFGLYREGYLSNKKIIGYARSKLDDKEFKDRIVGYFTKKNKGDEDKVQEFLLKCSYISAPYDKPDGGEKLNETINEFEKENNVEQSHRLFYALPPSVFIPVATEVKYVHPGSKGIARIIVEKPFHGDLQSAEELLNALKPIWKEELFRIDHYLGKEMVKNLLAFRFGNAFINASWDNRHISCIQJSFKEPFGTEGRGGYFDSIGIIRDVIQNHLLQVLTLLTMRPVSNDPEAVRDEKVRILKSISELDLNDLVGGYQKSEDEGKKPAYVDDETVPKSGKCVTFAAIGLHINTERWEGVPIILRAGKALNEGKVEIRVQYKQSTGFLNDIQRNELVIRVQPNEMMYMKNLSKVPVGSQKTTVTELDLYKDRYENFYIPEAYESLIRDAMKGDHNSFVRDDELIQSWKIFTPLLVHLEGPDAPAPEIYPYSGRGPASLTFLQDHDYFFESRDNYQWPVTRPDVLHK M
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#5c: Hypothetical protein PAS (525 AA) GI:238031215 Acc: CAY69138.1 77%	MLRTSPATKKALKSQINAFNVAALRFYSSLPQVPITLPNGKTYNQPTGLFINNEFVPSKQKTFAVLNPSTEEITHVYESREDVVELAVAAAQKAFDSTWSTQDPAERGKVLNKLADLIEHSETLAAIESLDNGKAISARGDVLVAVLKSCAGWADKVFGRVVTETGSSHFNYVRREPLGVCGQIIPWNFPLLMWSWKVGPALATGNTVVLKTAESTPLSALVSVQLVKEAGIPAGVHNIVSGFGKITGEAIATHPKIKKVAFTGSTATGRHIMKAAESNLKKVLTLEGGKSPNIVFNDANIKQAVANIILGIYNSGEVCCAGSRVYVQSGIYDELLAEFKTAENLVKGNPFDEDTFQGAQTSQQQLEKILGFVERGKKDGATLITGGGRLGDKGYFVQPTIFGDTVPEMEIVKEEIFGPVVTISKFDTIDEVVDLANDSQYGLAAGHSDINKVIDVAARIKSGTVWVNTYNDHFQHMVPFGFGGQSGIGREMGVEALENYTQYKAIRVKINHNE
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#6: Mitochondrial aldehyde dehydrogenase (501 AA) GI:238033249 Acc: CAY71271.1 81%	MTFAPPLEFEIDLPLNGLYTQPLGLFINNEFVEGVEGKLLPVINPCDETKITQVWEASAADVDRAVDAEADAFNNSVWATQDPLERGLMNNKLADLIDRDFNILAGIESIDNGKAYTSAQGDVTLAVNYIRSCAGWADKILGNVVDGNTHLNLVKREPLGVVGQIIPWNFPLLMWLKGLPALATGNTVVLKTAESTPLSGLVAKLIKEAGFPFVGNILSGFGNPAGAAIAHPRIKIAFTGSTATGRKIMEAAAKSNLKKVLTLEGGKSPNIVFEDADIQKTIHNIILGIFNNSGEVCCAGSRVYIQTIVYEEVLEAFKKETDNVKKVGGPFEEFGVQGPQTSLEQLNRILSYIKHGKDEGARVITGGSRVNRNGYIYKPTIFADVTEDMKIVKEEIFGPVVTITKFTSTDEVVGYANNTNYGLAAGIHTNNLNKAIDVASRIKAGVWVINTYNDFFHMMVPFGGYGESIGRELGAELDNYTQAKAIRIAYTP EHK
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#7a: Delta-aminolevulinate dehydratase (341 AA) GI:238033645 Acc: CAY71667.1 73%	MVHKAELYDDHPTQISSILSGGYNHPLLREWQHERQLNKNMFIFLFTVDRPDEEELPSLPNIKRFVGNKLIPYVGGGLVSKGLRAVILFGVPLKPGVKDEEGTAADDPEGPVIAIKHLRKNFPDLYIITDVCLCEYTSHGHCILYEDGTINRELSVRRIAIAVAVKYAQAGANSVAPSDMTDGRIRDIKEGLLSAGLAHKTFVMSYAAKFSGNLYGPFPRDAAGSCPSQGDRCYQLPSGGKGLAHRALRDMNGETDGIIVKPTFYLDIVADAYQLCKDYPICCYQVSGEYAMLHAAAEKNVLDLKSIAFEAHQGLFRAGARLIISYFTEFLEWLSE
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#7b: Mitochondrial alcohol dehydrogenase isozyme III (350 AA) GI:238031179 Acc: CAY69102.1 84%	MSPTIPTTKQKAVIFETNGGPLEYKDIPVPKPKSNELLINVKYSGVCHTDLHAWKGDWPLDNKLPLVGGHEGAGVVAAYGENVTGWEIGDYAGIKWLNGSCLNCEYCIQGAESSCAKADLSGFTHDGSFQQYATADATQAARIPKEADLAEVAPILCAGITVYKALKTADLRIGQVVAISGAGGGLSLAVQYAKALGLRVLGIDGGADKGEFVKSGLGAEVVFDTKT KDVVGAEPKLTNNGGPHGVINVSVPCHAINQSVQVVRTLGKVVVLVGLSGAVVNSDVFWHVLKSEIKGSYVGNREDSAEADLFRGLVKAPIKIIGLSELAKVYEQMEAGAIIGRYVVDTSK
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#7c: Malate dehydrogenase (342 AA) GI:238034064 Acc: CAY72086.1 77%	MVKVTVCGAAGGIGQLSLMFKNLPYVTTLALYDVVNVPVGKDLSHIDTDTKLESYLPENDGLEKALTGSDLVIIIPAGVPRKPGMTRDDLFAINAGIIRDLANGIAQFAPSFAVLVISNPVNSTVPIVAEILKNNVFNPKLFGVTTLDVCRANTFVAELSKDKEASAFDTRVLGGHSGETIVPVFSQSAPEVYKELSDQKAALVHRVQFGGDEVVKAKNAGSATSLSMAYAGYKLGHALLAIAINDTPNIIESTFVYLKDSKIKGAEEAFKYNELKLDSSSDVDFFALPVLVSSNGIEIKWDILEKVDAKETELLKIATGQLSKNIAGKTAFIAGN
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#7d: Putative protein, unknown function (328 AA) GI:238033788 Acc: CAY71810.1 77%	MVVAIEGGTGLGLMNLTKWPKPTPIDDAIETIRYAVEEAGVRYLNGGEFYNPLDSNLNLQYIQEFKARYPELYKKVLSLVKGAVSLVDVSPDSSPENLEKSINITHLPNNFLPIFEPARIDKRYSEETIKNLSKFVEDGRIGGISLSEVGADTIRRAAKVAPIACVEVEFSLTRDILHNGVLAACEDLNIPAIAYSPLGRGFLTGTINSKADIPEGDIRLSLERFNDDVEIHNKLVLHGLKKIADKKGVTLAQLSLAWLRKFGDKHVKVLPIPCSSPPRRVAENTKEISLTDSEFQEITDFAESVPIKGGRYNKASEAVLNG
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#8: Triose phosphate isomerase (248 AA) GI:238032989 Acc: CAY71012.1 73%	MARTFFVGGNFKMNGSKSIIHEIRLNNTKLPENVEVVIAPPYLPQAVTENKQKTVVYSAQNSFDKASGAYTGEVSVEALKDLGVPIYVILGHSERRTINKEDDAFIASKTFKALDQGLKVICIGETLEEKQANITLDVVKKQLQAVVDVSDWTNIVVAYEPVVAIGTGLAATPSDAQDVHKKIRDFLATVIGKDQAEKVRILYGGSVNGKNAVEFRDKADVDFLGVGGASLKPEFVDIINSRN
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#9a: Hypothetical protein (cyclophilin) PP7435 (161 AA) GI:328350030 Acc: CCA36430.1 66%	MTKTFDFVSSNDQPLGRIVFELYDDVPKTIENFRALCTGEKGYKDSIFHRVIPQFMLQGGDFTKFNGTGGKSIYGEKFADEFNIHKKHTKPLGLSMANAGPNTNGSQFFITTVPCPWLGDGHVVFGEVVDGLDVVSKIETLGSSSGATKTQLKINTSGEL
<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#9b: Cytosolic superoxide dismutase (154 AA) GI:238034030 Acc: CAY72052.1 94%	MVKAVAVLRGDSVGGTVVFEQSSSEPTTITYDIKGNSPNAERGFHIIHQFGDNTNGTSGAGPHFNPFGKTHGAPTDEARHVGD LGNVKTDAGVAKGVITDNQVKLIGETSLGRTVVIHGDGDDLGKGGHADSLKTGNAGGRPACGVIGLAA

<i>Komagataella phaffii</i> (<i>Pichia pastoris</i>)	#10: Mitochondria ATPase inhibitor (84 AA) GI:238029769 Acc: CAY67692.1 32%	MFQRTTATLVRQNKAIARFYSEGSTGAPRSDGSGDAFTKREKAQEDFYIKKHQAEQLAKLREQQKNQKEHLQNLKEINNIK EK
---	---	--

Table S2 Literature search results for soybean (*Glycine max*) leghemoglobin. The primary and most useful search was with PubMed, but additional searches were performed using Web of Science, Scopus and AGRICOLA. The abstracts and in some cases, whole papers were reviewed to understand relevance.

<i>Glycine max</i> only	<i>Glycine max</i> AND leghemoglobin	<i>Glycine max</i> AND allergen/allergy	<i>Glycine max</i> AND toxin/toxic
PubMed search results			
21,936 articles	201 articles When allergen was added to the search criteria, no reference came up.	381 articles for allergen and 665 articles for allergy. Further searches with leghemoglobin added to either search criteria yielded no results. One article was cited when searched with Soy AND allergy AND heme but was not relevant to the topic.	419 articles found for toxin and 342 articles found for toxic. An additional criterion of either leghemoglobin or heme was added to the search criteria with <i>Glycine max</i> AND toxic yielded 1 and 8 results, respectively. None were relevant to the topic.
Web of Science Core Collection search results			
19,230 articles	90 articles When allergen was added to the search criteria, no reference came up.	103 articlesnd for allergen and 35 articles for allergy. Further searches with leghemoglobin added to either search criteria yielded no results.	67 articles found for toxin and 276 articles found for toxic. An additional criterion of either leghemoglobin or heme was added to the search criteria with <i>Glycine max</i> AND toxic yielded 2 and 2 results, respectively. None were relevant to the topic.
Scopus search results			
48,222 articles	143 articles When allergen was added to the search criteria, no reference came up.	319 articles for allergen and 283 articles for allergy. Further searches with leghemoglobin added to either search criteria yielded no results.	288 articles found for toxin and 679 articles found for toxic. An additional criterion of either leghemoglobin or heme was added to the search criteria with <i>Glycine max</i> AND toxic yielded 3 and 3 results, respectively. None were relevant to the topic.
National Agricultural Library Catalog (AGRICOLA) search results			
Over 10,000 articles	123 articles When allergen was added to the search criteria, no reference came up.	31 articles for allergen and 6 articles for allergy. Further searches with leghemoglobin added to either search criteria yielded no results.	26 articles found for toxin and 135 articles found for toxic. An additional criterion of either leghemoglobin or heme was added to the search criteria with <i>Glycine max</i> AND toxic yielded 2 and 1 results, respectively. None were relevant to the topic.

Table S3 Literature search results for the recombinant yeast host *Pichia pastoris* (synonym *Komagataella phaffii*). Searches with PubMed were most productive. Additional searches were performed using Web of Science, Scopus and AGRICOLA. The abstracts and in some cases whole papers were reviewed to understand the relevance.

<i>Pichia pastoris</i> or <i>Komagataella phaffii</i>	<i>Pichia pastoris</i> AND allergen	<i>Pichia pastoris</i> AND toxin	Individual endogenous (17) <i>Pichia pastoris</i> proteins <i>Pichia pastoris</i> AND individual protein name
PubMed search results			
5,243 articles for <i>Pichia pastoris</i> . 6 articles for <i>Komagataella phaffii</i> . None of the publications on <i>K. phaffii</i> were related to allergens or toxins of the organism.	128 articles, most relate to expression of heterologous genes and proteins. 5 articles when NOT recombinant was added.	185 articles 26 articles when NOT recombinant was added. Carefully reading all the abstracts demonstrated that only one is relevant to topics, which demonstrated that no toxins active against pathogens or other yeasts were identified in <i>Pichia pastoris</i> [1].	Searches with the following individual identified proteins did NOT identify any publications that described allergenicity or toxicity. Alpha-aminoadipate reductase: 0 article Cobalamin-independent methionine synthase: 0 articles Aconitase: 2 articles Transketolase: 5 articles Glycerol kinase: 5 articles Catalase A: 21 articles, 0 article when AND Allerg* added GAPDH: 1 article Hypothetical protein PAS: 0 article Mitochondrial aldehyde dehydrogenase: 1 article Delta-aminovulinate dehydrogenase: 1 article Mitochondrial alcohol dehydrogenase: 2 articles Malate dehydrogenase: 1 article Unknown function (possibly pyridoxal reductase: 0 Triosphosphate isomerase: 0 Cyclophilin: 0 Cytosolic superoxide dismutase: 3 articles Mitochondria ATPase inhibitor: 0
Web of Science Core Collection search results			
8,399 articles for <i>Pichia pastoris</i> . 18 articles for <i>Komagataella phaffii</i> . None of the publications on <i>K. phaffii</i> were related to allergens or toxins of the organism.	188 articles, most relate to expression of heterologous genes and proteins. 36 articles when NOT recombinant was added.	172 articles 55 articles when NOT recombinant was added. The same paper was identified [1].	Searches with the following individual identified proteins did NOT identify any publications that described allergenicity or toxicity. Alpha-aminoadipate reductase: 0 article Cobalamin-independent methionine synthase: 0 article Aconitase: 1 articles Transketolase: 3 articles Glycerol kinase: 10 articles Catalase A: 49 articles, 0 article when AND Allerg* added GAPDH: 8 articles Hypothetical protein PAS: 0 article Mitochondrial aldehyde dehydrogenase: 0 article Delta-aminovulinate dehydrogenase: 0 article Mitochondrial alcohol dehydrogenase: 0 article Malate dehydrogenase: 0 article Unknown function (possibly pyridoxal reductase): 0 article Triosphosphate isomerase: 0 article Cyclophilin: 0 article Cytosolic superoxide dismutase: 0 article Mitochondria ATPase inhibitor: 0 article
Scopus search results			
7,378 articles for <i>Pichia pastoris</i> .	137 articles, most relate to expression of heterologous genes	121 articles 37 articles when NOT	Searches with the following individual identified proteins did NOT identify any publications that described allergenicity or toxicity.

20 articles for <i>Komagataella phaffii</i> . None of the publications on <i>K. phaffii</i> were related to allergens or toxins of the organism.	and proteins. 42 articles when NOT recombinant was added.	recombinant was added. One paper [2] investigated <i>Tetrapispora phaffii</i> killer toxin Kpkt which is heterologously produced in <i>Phicia pastoris</i> . It is not an endogenous protein from <i>Pichia</i> . Besides, Kpkt is known to be active to spoilage yeast in wine-making [3]. Thus, Scopus search failed to find any evidence of toxin existing in <i>Pichia</i> .	Alpha-aminoadipate reductase: 0 article Cobalamin-independent methionine synthase: 0 article Aconitase: 0 article Transketolase: 7 articles Glycerol kinase: 15 articles Catalase A: 48 articles, 0 article when AND Allerg* added GAPDH: 9 articles Hypothetical protein PAS: 1 article Mitochondrial aldehyde dehydrogenase: 1 article Delta-aminovulinate dehydrogenase: 0 article Mitochondrial alcohol dehydrogenase: 5 articles Malate dehydrogenase: 6 articles Unknown function (possibly pyridoxal reductase): 0 article Triosphosphate isomerase: 0 article Cyclophilin: 2 articles Cytosolic superoxide dismutase: 3 articles Mitochondria ATPase inhibitor: 0 article
National Agricultural Library Catalog (AGRICOLA) search results			
1,971 articles for <i>Pichia pastoris</i> . 7 articles for <i>Komagataella phaffii</i> . None of the publications on <i>K. phaffii</i> were related to allergens or toxins of the organism.	13 articles, most relate to expression of heterologous genes and proteins. 5 articles when NOT recombinant was added.	24 articles 8 articles when NOT recombinant was added.	Searches with the following individual identified proteins did NOT identify any publications that described allergenicity or toxicity. Alpha-aminoadipate reductase: 0 article Cobalamin-independent methionine synthase: 0 article Aconitase: 0 article Transketolase: 0 article Glycerol kinase: 2 articles Catalase A: 7 articles GAPDH: 4 articles Hypothetical protein PAS: 0 article Mitochondrial aldehyde dehydrogenase: 0 article Delta-aminovulinate dehydrogenase: 0 article Mitochondrial alcohol dehydrogenase: 1 article Malate dehydrogenase: 0 article Unknown function (possibly pyridoxal reductase): 0 article Triosphosphate isomerase: 0 article Cyclophilin: 0 article Cytosolic superoxide dismutase: 3 articles Mitochondria ATPase inhibitor: 0 article

Table S4 Overall FASTA3 search of AllergenOnline version 2016 (AOL v16) with the leghemoglobin. Summary results from FASTA3 full-length search of AllergenOnline.org database with the AA sequence of LegHb. Results only identified modest sequence matches to the occupational allergen, hemoglobin from larvae of the *Chironomus thummi*. that are sometimes called blood worms due to high hemoglobin content.

Quary Protein	Sequence GI# [Acc#]	Organism	Description	Length aa	E score	Identity (%)	aa Alignment length
leghemoglobin (LegHb) GI:126241 <i>Glycine max</i>	56405052 [P84296.1]	<i>Chironomus thummi thummi</i>	Globin (Insect hemoglobin)	161	0.00024	26.2	145
	121244 [P12548.1]	<i>Chironomus thummi thummi</i>	Globin	161	0.00043	25.9	147
	56405054 [P84298.1]	<i>Chironomus thummi thummi</i>	Globin	161	0.0009	25.0	144
	121248 [P12549.1]	<i>Chironomus thummi thummi</i>	Globin	161	0.0012	25.0	144
	121249 [P12550.1]	<i>Chironomus thummi thummi</i>	Globin	162	0.0012	25.2	151
	2506460 [P02221.2]	<i>Chironomus thummi thummi</i>	Globin	158	0.0039	24.2	149
	1707908 [P02222.2]	<i>Chironomus thummi thummi</i>	Globin	160	0.0053	25.2	139
	121259 [P02228.1]	<i>Chironomus thummi thummi</i>	Globin	151	0.0068	26.9	119
	121237 [P02227.1]	<i>Chironomus thummi thummi</i>	Globin	151	0.019	26.3	137
	2506461 [P02224.2]	<i>Chironomus thummi thummi</i>	Globin	162	0.049	24.0	150
	1707911 [P02223.2]	<i>Chironomus thummi thummi</i>	Globin	161	0.057	22.9	140
	121256 [P02231.1]	<i>Chironomus thummi thummi</i>	Globin	151	0.32	22.1	149

Table S5 BLASTP of NCBI Entrez with leghemoglobin using Entrez query keyword “allergen”, “allergy”, “toxic” and “toxin”. BLAST with keyword “allergy” gave a list of 15 protein sequences on 27th February, 2016 and the top 2 were shown in the table. All are poorly scoring identity –sequence length matches with large *E* scores, suggesting low level homology.

Keyword	BLAST results						
	Sequence Accession #	Organism	Description	Length aa	<i>E</i> score	Identity (%)	Query Cover (%)
allergen	No matches found						
allergy	EOQ89444.1	<i>Leptospira yanagawae</i> serovar <i>Saopaulo</i>	adenylate/guanylate cyclase catalytic domain protein	440	7e-05	29	93
	EJO56091.1	<i>Burkholderia multivorans</i>	nitric oxide dioxygenase	403	0.002	43	31
toxic	No matches found						
toxin	BAO69646.1	<i>Bordetella bronchiseptica</i>	nitric oxide dioxygenase	402	0.027	35	31
	ALX21654.1	<i>Bordetella pertussis</i>	dihydropteridine reductase	402	0.028	35	31

Table S6 Bioinformatics Summary Results of Protein Band #1. The 1400 AA protein, Alpha

aminoadipate reductase, GI: 238030060 was compared with AOL v 16 and with NCBI Protein on 17th

May, 2016.

AOL full FASTA3	No match to any protein with an <i>E</i> score of < 1 was found. Four low identity short segment matches with <i>E</i> scores between 1 and 10. The proteins with short-low identity matches included icarapin from honey bee, ovotransferrin from chicken egg and Ani s 7 from <i>Anisakis simplex</i> .
AOL 80mer FASTA3	No match to any protein with identity of >35% over any 80 AA segment.
AOL Exact 8 AA	No exact 8mer match with any protein.
NCBI BLASTP Allergen	Two low scoring identity matches to non-homologous, hypothetical proteins with <i>E</i> scores >1 e-05, and alignments of < 25% of the length and < 30% identity.
NCBI BLASTP Toxin or Toxic	A number of low scoring identity matches to homologous proteins (small <i>E</i> scores) were found that are proteins from organisms that are classified as toxic to insects (e.g. <i>Saccharomyces</i>), but no significant alignment to an obvious toxin or toxic proteins.
NCBI BLASTP No keyword	Many matches of high scoring identity with protein homologous proteins from other yeasts and molds. The highest scoring match to a <i>Saccharomyces sp.</i> protein has an <i>E</i> score = 0.0, and 60% identity in a 1418 AA alignment.
Conclusion: There is a very low risk of Allergy or toxicity identified for protein #1 based on lack of sequence matches to any known allergens. Therefore, risks from consumption of this protein appear to be minimal.	

Table S7 Bioinformatics Summary Results of Protein Band #2a. The 768 AA Cobalamin-independent methionine synthase protein, GI: 238030843 was compared with AOL v16 and with NCBI Protein on

17th May, 2016.

AOL full FASTA3	One significant match was identified to Sal k 3 from Russian thistle (<i>Salsola kali</i>). Sal k 3 was first identified as a pollen allergen in 2011 [4]. While it is not clear how important Sal k 3 is an airway allergen, there hasn't been any publication of allergy upon digestion, and no cross-reactivity between Sal k 3 and any food allergen has been reported. In addition, a BLASTP with Sal k 3 revealed highly homologous proteins from wide food sources that are not known as allergens with much higher identities to Sal k 3 than found with band #2a (Figure 2a). The lack of reported cross-reactivity between Sal k 3 and other common proteins and sources indicate it is unlikely that protein #2a will be a cross-reactive issue.
AOL 80mer FASTA3	The only protein with an 80mer match was Sal k 3, with 77.5% identity as the highest alignment.
AOL Exact 8 AA	The only protein with an exact 8mer match is Sal k 3.
NCBI BLASTP Allergen	Two low scoring identity matches to non-homologous, hypothetical proteins with <i>E</i> scores >1 e-05, and alignments of < 25% of the length and < 30% identity.
NCBI BLASTP Toxin or Toxic	A number of low scoring identity matches to homologous proteins (small <i>E</i> scores) that are proteins from organisms that are classified as toxic to insects (e.g. <i>Bacillus thuringiensis</i>), but no significant alignment to an obvious toxin or toxic proteins.
NCBI BLASTP No keyword	Many matches of high scoring identity with homologous proteins from other yeasts and molds. The highest scoring match to a <i>Saccharomyces sp.</i> protein has an <i>E</i> score = 0.0, and 77% identity in a 765 AA alignment.
Conclusion: There is a minimal risk of allergy based from the modest alignment with Sal k 3. There was no significant match to a toxin, so the likely risk of toxicity is extremely low. Therefore, risks from consumption of this protein appear to be minimal.	

Table S8 Bioinformatics Summary Results of Protein Band #2b. The 780 AA Aconitase protein,

GI:254564667 was compared with AOL v16 and with NCBI Protein on 17th May, 2016.

AOL full FASTA3	There were two alignments with <i>E</i> scores between 0.1 and 1 (<i>Aspergillus sp.</i> endo-chitosane 238 AA, 31% identity over 80 AA; <i>Sarcoptes scabiei</i> inactive cysteine protease, 340 AA, 25% identity over 139 AA), and 10 alignments with <i>E</i> scores between 4 and 10, various proteases with low identity scores (24% to 37% identity).
AOL 80mer FASTA3	No matches of >35% identity over 80 AA with any protein in AOL v16.
AOL Exact 8 AA	No exact 8 AA match with any protein in AOL v16.
NCBI BLASTP Allergen	One very low scoring alignment with an alkyl hydroperoxide annotated as a “Mal allergen” (no reference) from an Antarctic coastal bacteria, genomic sequence. The <i>E</i> score was 0.86, 35% identity over 72 AA and three other irrelevant lower identity alignments.
NCBI BLASTP Toxin or Toxic	A number of statistically significant alignments with <i>E</i> scores smaller than 1e-50, one with 0.0 were found. However most of the alignments were modest identities of 50% or less, indicating homology. Most of the source organisms were toxic, but no obvious reference to toxicity of the proteins. There was one higher scoring alignment of 66% with <i>E</i> score 0.0, to a bovine homologue of the aconitase enzyme (GI: 157831069, bovine 4-hydroxy-trans-aconitate) can bind to a “toxic” substrate, fluorocitrate. The protein itself is not toxic. And the fluorocitrate and fluoroacetate inhibits the enzyme, causing toxicity. The LegHb and plant based meats do not contain fluoroacetate or flurocitrate and the aconitase protein is not a functional part of food.
NCBI BLASTP No keyword	Many matches of high scoring identity with protein homologous proteins from other yeasts and molds. The highest scoring match to a <i>Saccharomyces sp.</i> protein has an <i>E</i> score = 0.0, and 81% identity in a 669 AA alignment.
Conclusion: There is a very little risk of allergy based on the modest alignment with endo-chitosane and inactive protease. There were moderate alignments to proteins from toxic organisms, but only as common homologues. Thus, the likely risk of toxicity is extremely low. Therefore, risks from consumption of this protein appear to be minimal.	

Table S9 Bioinformatics Summary Results of Protein Band #3. The 679 AA Transketolase protein,

GI:238030057 was compared with AOL v16 and with NCBI Protein on 17th May, 2016.

AOL full FASTA3	No matches to proteins in AOL v16 with an <i>E</i> score <1. There were >20 matches with <i>E</i> score between 1 and 10 with low identities of <35% and they were generally short, 49 to 134 AA. These matches were not significant.
AOL 80mer FASTA3	No identity match of >35% over any segment of 80 amino acids compared to any protein.
AOL Exact 8 AA	No exact 8 AA matches with any protein.
NCBI BLASTP Allergen	No significant alignment with any protein in NCBI with using the keyword allergen.
NCBI BLASTP Toxin or Toxic	Many small <i>E</i> score identity matches of just under 50% identity to transketolases of bacteria (e.g. <i>Bacillus cereus</i>) or cyanobacteria that are known to be toxic organisms. These are homologues of the enzyme involved in sugar catabolism from photosynthesis. No references were obvious showing any toxic effects of these enzymes on consumers.
NCBI BLASTP No keyword	Many matches of high scoring identity with protein homologous proteins from other yeasts and molds. A transketolase from the commonly used food yeast, <i>Saccharomyces cerevisiae</i> has an <i>E</i> score of 0 and 70% overall identity to the transketolase of <i>P. pastoris</i> .
Conclusion: There is no identified risk of allergy or toxicity from the transketolase protein. The higher scoring alignment with transketolase of <i>Saccharomyces cerevisiae</i> provides additional assurance of safety. Therefore, risks from consumption of this protein appear to be minimal.	

Table S10 Bioinformatics Summary Results of Protein Band #4. The 621 AA Glycerol kinase protein,

GI:238034027 was compared with AOL v16 and with NCBI Protein on 17th May, 2016.

AOL full FASTA3	No matches to proteins in AOL v16 with an <i>E</i> score <1. There were >20 matches with <i>E</i> score between 1 and 10 with low identities of <35% and they were generally short, 49 to 134 AA. These matches were not significant.
AOL 80mer FASTA3	No matches of >35% identity over 80 AA were found compared to any protein in AOL v16.
AOL Exact 8 AA	No exact 8 AA matches were identified to any protein in AOL v16.
NCBI BLASTP Allergen	No significant alignments were identified.
NCBI BLASTP Toxin or Toxic	A number of low scoring identity matches to homologous proteins (small <i>E</i> scores e.g. 6e-118) that are proteins from organisms that are classified as toxic to insects (e.g. <i>Bacillus thuringiensis</i>), or mammals (e.g. <i>Mycobacterium tuberculosis</i>), with up to 41% identity matches and many gaps over 550 AA, but no significant alignment to an obvious toxin or toxic proteins.
NCBI BLASTP No keyword	Many matches of high scoring identity with protein homologous proteins from other yeasts and molds were found. The highest scoring alignment with a <i>Saccharomyces sp.</i> was with an <i>E</i> score of 0.0 and identity of 53% over 664 AA.
Conclusion: There does not seem to be a risk of allergy based on the lack of alignment to allergens. There are modest alignments to proteins that have not been demonstrated to be toxins, but are from organisms that are known to be toxic. However, with higher identities to proteins from other yeasts, including <i>Saccharomyces sp.</i> , it is very unlikely that this protein is a toxin. Therefore, risks from consumption of this protein appear to be minimal.	

Table S11 Bioinformatics Summary Results of Protein Band #5a. The 510 AA Catalase A protein,

GI:254569930 was compared with AOL v16 and with NCBI Protein on 17th May, 2016.

AOL full FASTA3	One highly significant match with 2.6 e-058 but only 37% identity over 475 AA was identified to a putative allergen Pen c 30, a catalase of <i>Penicillium citrinum</i> , identified as an allergen based on IgE binding to mold allergic subjects in Taiwan [6]. There were two Asn-linked glycans on the natural <i>Penicillium citrinum</i> protein and the deglycosylated form had reduced binding. There was no published allergic reaction to ingestion of Pen c 30. <i>Penicillium camemberti</i> is widely consumed with brie and other soft-ripened cheeses [7], and <i>Aspergillus oryzae</i> is widely used in Asia as koji in soy sauce and soy paste fermentation [8], both of which have catalase that are closer to the allergenic ones than is the Pichia protein (Figure 2b). Therefore, cross-reactivity between #5a and Pen c 30 is highly unlikely.
AOL 80mer FASTA3	The only alignment with >35% identity over 80 AA was to Pen c 30 and the best aligned 80 AA segment was 60% identity, indicating restricted conservation of sequence to a subsection of the protein.
AOL Exact 8 AA	No exact 8 AA matches were identified to any protein, not even to Pen c 30.
NCBI BLASTP Allergen	Only the single alignment of Pen c 30, with an <i>E</i> score of 9e-93 and 38% identity over 424 AA.
NCBI BLASTP Toxin or Toxic	Many significant (very small <i>E</i> score) alignments were identified (e.g. 5e-149 for toxin, 0.0 for toxic) to protein homologues from organisms noted for toxins or toxicity (e.g. <i>Bacillus sp.</i> , <i>Enterococcus faecalis</i> , <i>Streptomyces sp.</i> , <i>Clostridium sp.</i>). The alignments were approximately 50% for proteins identified with “toxin”, and 71% as the highest for those identified with “toxic” organisms (<i>Candida boidinii</i>). The enzyme is common to all organisms living in oxygen containing environments as it detoxifies hydrogen peroxides by converting them to H ₂ O and O ₂ . It is not toxic to mammalian cells. No published evidence was found that catalase A is a toxin.
NCBI BLASTP No keyword	Many high scoring identity matches were found with homologous proteins from other yeasts and molds. The full-length of #5a represents a conserved catalase A common to most yeast and fungi. The highest scoring alignment with a <i>Saccharomyces sp.</i> was with an <i>E</i> score of 0.0 and identity of 66% over 494 AA.
Conclusion: There is a minimal risk of allergy based on a relatively poor alignment to the allergen Pen c 30. There are much higher alignments to other proteins including the catalase A proteins of a variety of yeasts and molds including from <i>Saccharomyces sp.</i> There are modest alignments to proteins that have not been demonstrated to be toxins, but are from organisms that are known to be toxic. Therefore, risks from consumption of this protein appear to be minimal.	

Table S12 Bioinformatics Summary Results of Protein Band #5b. The 504 AA glucose 6 phosphate

dehydrogenase protein (G6PD), GI: 238031000 was compared with AOL v16 and with NCBI Protein on 17th

May, 2016.

AOL full FASTA3	Low scoring matches were identified with <i>E</i> scores ranging from 0.91 to 9.4 for 6 proteins that had identity matches of 24% to 47% over short alignments of 88 down to 30 AA. These proteins were not homologues and the identity matches appear random.
AOL 80mer FASTA3	The only alignment with >35% identity over 80 AA was to the putative allergen, <i>Blattella germanica</i> , German cockroach, Bla g 3 (hemocyanin) with 37% identity as the best aligned 80 AA segment. Bla g 3 only had 25% identity over full alignment with #5b protein with a highly insignificant <i>E</i> score of 33.
AOL Exact 8 AA	No exact 8 AA matches were identified to any protein in AOL v16.
NCBI BLASTP Allergen	No significant identity match was identified in NCBI with keyword allergen.
NCBI BLASTP Toxin or Toxic	Many significant (very small <i>E</i> score) alignments were identified (as low as 6e-93 for toxin, 6e-95 for toxic) to homologues from organisms noted for toxins or toxicity (e.g. <i>Bacillus sp.</i> , <i>Enterococcus faecalis</i> , <i>Streptomyces sp.</i> , <i>Clostridium sp.</i>). The alignments were approximately 35% for proteins identified with "toxin", and 35% as the highest for those identified with "toxic" organisms (<i>Fictibacillus phosphorivorans</i>). The enzyme is common to all organisms. It is not a toxin and is not toxic to mammalian cells.
NCBI BLASTP No keyword	Many high scoring identity matches were found with homologous proteins from other yeasts and molds. The full-length of #5b alignment to <i>Saccharomyces sp.</i> was with an <i>E</i> score of 0.0 and identity of 64% over 495 AA.
Conclusion: There does not seem to be a risk of allergy based on the lack of alignment to allergens beyond the low-level alignment to Bla g 3. Many homologues were identified to proteins from yeasts, including <i>Saccharomyces sp.</i> There are modest alignments to proteins that have not been demonstrated to be toxins, but are from organisms that are known to be toxic. Therefore, risks from consumption of this protein appear to be minimal.	

Table S13 Bioinformatics Summary Results of Protein Band #5c. The 525 AA hypothetical protein

(PAS), GI: 238031215 was compared with AOL v16 and with NCBI Protein on 17th May, 2016.

AOL full FASTA3	Two very significant scoring alignments were identified, one with an <i>E</i> score of 5e-126 and 58% identity with Cla h 10 (<i>Davidiella sp.</i> aldehyde dehydrogenase) and one with 1e-124 <i>E</i> score and 58% identity to Alt a 10 (<i>Alternaria alternata</i> aldehyde dehydrogenase). Those proteins are minor hypothetically airway allergens identified only by low-level IgE binding in the studies used to identify them as putative allergens. The next highest alignments were to diverse species and of much greater <i>E</i> scores and lower identities.
AOL 80mer FASTA3	Three alignments were found with >35% identity. The highest scoring ones were to Cla h 10 and Alt a 10 with highest scoring 80mers of 72% identity. The third alignment was a storage mite, <i>Lepidoglyphus destructor</i> Lep d 13 with best identity of 35%, which only had 29.6% identity over full alignment with #5c protein with a highly insignificant <i>E</i> value of 96.
AOL Exact 8 AA	A number of exact 8 AA matches were identified to the Cla h 10 protein and Alt a 10 protein.
NCBI BLASTP Allergen	There were a few significant alignments using NCBI with keyword allergen. The best scoring identities were 58-59% over full length to Cla h 10 and Alt a 10, two mold allergens.
NCBI BLASTP Toxin or Toxic	Some significant (very small <i>E</i> score) alignments were identified (as low as 2e-171 for toxin, 2e-177 for toxic) to homologues from organisms noted for toxins or toxicity (e.g. <i>Bacillus sp.</i>). The alignments were approximately 53% for proteins identified with “toxin”, and “toxic” organisms. The enzyme is common to all organisms. It is not a toxin and is not toxic to mammalian cells.
NCBI BLASTP No keyword	Many high scoring identity matches were found with homologous proteins from other yeasts and molds. The full-length PAS alignment to <i>Saccharomyces sp.</i> was with an <i>E</i> score of 0.0 and identity of 69% over 512 AA.
Conclusion: There does not seem to be a risk of allergy based on the lack of alignment to allergens. There are modest alignments to proteins that have not been demonstrated to be toxins, but are from organisms that are known to be toxic. Many homologues were identified to proteins from yeasts, including <i>Saccharomyces sp.</i> Therefore, risks from consumption of this protein appear to be minimal.	

Table S14 Bioinformatics Summary Results of Protein Band #6. The 501 AA mitochondrial aldehyde dehydrogenase protein, GI: 238033249 was compared with AOL v16 and with NCBI Protein on 17th May, 2016.

AOL full FASTA3	Cla h 10 and Alt a 10 were the two matches with <i>E</i> scores of 2.4e-138 and 1.2e-131 and identities of 60% and 57%. They are minor hypothetically airway allergens identified only by low-level IgE binding in the studies used to identify them as putative allergens [5].
AOL 80mer FASTA3	Cla h 10 and Alt a 10 were identified in alignments with best-scoring identities of 76% and 74%.
AOL Exact 8 AA	Cla h 10 and Alt a 10 had a large number of 8 AA matches as expected based on high identity matches.
NCBI BLASTP Allergen	High scoring mitochondrial aldehyde dehydrogenase fungal proteins were identified in NCBI as the clear alignments using keyword allergen. <i>Cladosporium (Davidiella sp.)</i> and <i>Aureobasidium namibiae</i> were highest scoring with <i>E</i> scores of 0.0 and identities of 61% and 60%. These proteins are putative allergens, IgE binding, no biological activity (<i>e.g.</i> BHR, SPT).
NCBI BLASTP Toxin or Toxic	Slightly lower scoring alignments (compared to the fungal mitochondrial aldehyde dehydrogenases) or 6e-170 and 4e-165 and identity scores of 53% and 51% respectively were found for toxin. Interestingly, a human mitochondrial aldehyde dehydrogenase was the highest scoring protein with “toxic” as a search term, having an <i>E</i> score of 6e-173 and 52% identity. The next scoring match was to aldehyde dehydrogenase of <i>Bacillus thuringiensis</i> were identified using toxin and toxic keywords. The matches were to homologous proteins from bacteria (<i>e.g. Bacillus thuringiensis</i>) that do not have mitochondria, but the function is still needed in these bacteria. There is no published evidence was found that any aldehyde dehydrogenase is a toxin.
NCBI BLASTP No keyword	Many matches of high scoring identity with homologous proteins from other yeasts and molds were identified. The highest scoring alignment with a <i>Saccharomyces sp.</i> was with an <i>E</i> score of 0.0 and identity of 62% over 488 AA.
Conclusion: There is a very minor potential risk of allergy based on the identity alignments to two putative mold airway allergens. The enzyme is ubiquitous and there has not been proof that these proteins cause allergic disease, only that they can bind IgE from some airway sensitized individuals. There is a slightly better alignment to the homologous protein from <i>Saccharomyces sp.</i> Therefore, risks from consumption of this protein appear to be minimal.	

Table S15 Bioinformatics Summary Results of Protein Band #7a. The 341 AA Delta-aminolevulinate dehydratase protein, GI: 238033645 was compared with AOL v16 and with NCBI Protein on 17th May, 2016.

AOL full FASTA3	Very low-scoring alignments with <i>E</i> scores of 4 to 10 were identified to glucanase enzymes from diverse sources (e.g. rubber tree, grass pollen, cockroach and wheat). These alignments were not significant.
AOL 80mer FASTA3	No matches were identified with >35% identity in any 80 AA match with AOL v16.
AOL Exact 8 AA	No exact 8 AA matches were identified with AOL v16.
NCBI BLASTP Allergen	Three short, very minor identity matches with protein segments from <i>Parasteatoda tepidariorum</i> , not a known allergen.
NCBI BLASTP Toxin or Toxic	Modest scoring alignments were identified using toxin and toxic keywords. The matches of approximately 40% identity were to homologous proteins from bacteria. No published evidence was found that these proteins are toxins.
NCBI BLASTP No keyword	Many matches of high scoring identity with homologous proteins from other yeasts and molds were identified. The highest scoring alignment with a <i>Saccharomyces sp.</i> was with an <i>E</i> score of 0.0 and identity of 76% over 340 AA.
Conclusion: It is unlikely that there is any risk of allergy for this protein. The same is true for toxicity. In addition, there are very good alignments to homologous proteins from many molds including <i>Saccharomyces sp.</i> , with 76% identity over 340 AA. Therefore, risks from consumption of this protein appear to be minimal.	

Table S16 Bioinformatics Summary Results of Protein Band #7b. The 350 AA mitochondrial alcohol dehydrogenase protein, GI: 238031179 was compared with AOL v16 and with NCBI Protein on 17th May, 2016.

AOL full FASTA3	One significant-scoring alignment was identified with an <i>E</i> score of 1.2e-108 and 76% identity over the full-length (350AA) to <i>Candida albicans</i> Can a 1.0101. This protein should be considered a putative allergen as one publication demonstrated IgE binding to the apparent band of a partially purified protein [9] and was also expressed as a recombinant protein with very light IgE binding (compared to the native) by a pool of 4 sera [10], which had not absolutely demonstrated that the large number of subjects bound IgE to this specific protein. There was no demonstration of biological activity.
AOL 80mer FASTA3	Only the single protein Can a 1.0101 identified by full-FASTA was identified with >35% identity in any 80 AA match with AOL v16.
AOL Exact 8 AA	Only Can a 1.0101 contained exact 8 AA matches compared to protein band #7b.
NCBI BLASTP Allergen	Two high identity matches (<i>E</i> scores 0.0, identities of 80% and 76%) were found to a dehydrogenase of <i>Candida albicans</i> as a putative allergen. A third alignment was identified with an <i>E</i> score of 0.001 and identity of 28% over 120 AA to <i>Salmonella enterica</i> putative oxidoreductase.
NCBI BLASTP Toxin or Toxic	Modest scoring alignments were identified using toxin and toxic keywords to alcohol dehydrogenase enzymes of bacteria that are associated with toxicity (<i>Corynebacterium ulcerans</i> , <i>Bacillus thuringiensis</i> , <i>B. cereus</i> , <i>Escherichia coli</i> and <i>Streptococcus sp.</i>). The <i>E</i> scores were small 1e-87, 2e-55 and percent identities 36% to 42%. However, no published evidence was found that these proteins are toxins.
NCBI BLASTP No keyword	Many matches of high scoring identity with homologous proteins from other yeasts and molds were identified. The highest scoring alignment with a <i>Saccharomyces sp.</i> was with an <i>E</i> score of 0.0 and identity of 74% over 347 AA.
Conclusion: It is unlikely that there is any risk of allergy or toxicity for this protein. In addition, there are very good alignments to homologous proteins from many molds including <i>Saccharomyces sp.</i> Therefore, risks from consumption of this protein appear to be minimal.	

Table S17 Bioinformatics Summary Results of Protein Band #7c. The 342 AA malate dehydrogenase,

GI:238034064 protein was compared with AOL v16 and with NCBI Protein on 17th May, 2016.

AOL full FASTA3	<i>Malassezia furfur</i> Mala f 4.0101 was identified with an <i>E</i> score of 4.2e-54 and 51% identity over the full-length (341AA). Mala f 4 should be considered a putative allergen as the publication demonstrating IgE binding to a natural, semi-purified protein, but there was another isoform isolated and two other proteins from the partially purified Mala f 4 [11]. There was no demonstration of biological activity of this protein.
AOL 80mer FASTA3	Two proteins were identified with >35% identities. One was as expected, Mala f 4.0101, with a highest identity score of 70%. The second was to Pis s 2, the convicilin of <i>Pisum sativum</i> with 36.2% identity as the best match. However, Pis s 2 only had 31% identity with <i>E</i> score of 45 to protein #7c over full FASTA3.
AOL Exact 8 AA	Only Mala f 4.0101 contained exact 8 AA matches compared to protein band #7c with AOL v16.
NCBI BLASTP Allergen	The only match by BLASTP was to Mala f 4.0101 with an <i>E</i> score of 6e-71 and an identity of 45%.
NCBI BLASTP Toxin or Toxic	Significant scoring alignments were identified using toxin and toxic keywords to the malate dehydrogenase enzyme. The highest scoring hit was to a rat (<i>Rattus norvegicus</i>) with <i>E</i> score 2e-94 and 47% identity. Then a number of toxic organisms (<i>Vibrio cholera</i> and <i>E. coli</i>) with slightly lower identities and slightly higher <i>E</i> scores.
NCBI BLASTP No keyword	Many matches of high scoring identity with homologous proteins from other yeasts and molds were identified. The highest scoring alignment with a <i>Saccharomyces sp.</i> was with an <i>E</i> score of 2e-115 and identity of 57% over 346 AA.
Conclusion: It is unlikely that there is any risk of allergy for this protein as only one putative dermal allergen had a significant match. The risk of toxicity is similarly low. In addition, there are very good alignments to homologous proteins from many molds including <i>Saccharomyces sp.</i> Therefore, risks from consumption of this protein appear to be minimal.	

Table S18 Bioinformatics Summary Results of Protein Band #7d. The 328 AA putative protein,

GI:238033788 was compared with AOL v16 and with NCBI Protein on 17th May, 2016.

AOL full FASTA3	One low-scoring alignment was identified with an <i>E</i> score of 0.23 and 24.6% identity over the full-length (328AA) to <i>Juniperus occidentalis</i> Jun o 4.0101. This alignment is very low scoring and unlikely to be relevant to risks. Besides, there was no demonstration of biological activity.
AOL 80mer FASTA3	No alignments with >35% identity over any 80 AA match were identified.
AOL Exact 8 AA	No exact 8 AA match was identified to protein band #7d.
NCBI BLASTP Allergen	No significant match was identified using BLASTP and keyword Allergen.
NCBI BLASTP Toxin or Toxic	Modest scoring alignments were identified using toxin and toxic keywords to the malate dehydrogenase enzyme. The highest scoring hit was to a bacterial protein (<i>Escherichia coli</i> aldehyde oxidase) with <i>E</i> score 2e-39 and 32% identity. Then a number of toxic organisms (<i>E. coli</i> , <i>Microcystis aeruginosa</i> , <i>Raphidiopsis brookii</i>) with slightly lower identities and slightly higher <i>E</i> scores.
NCBI BLASTP No keyword	Many matches of very high scoring identities with homologous proteins from other yeasts and molds were identified with identities from 42% to 86%.
Conclusion: It is unlikely that there is any risk of allergy for this protein. The risk of toxicity is similarly low. In addition, there are very good alignments to homologous proteins from many molds including <i>Saccharomyces sp.</i> Therefore, risks from consumption of this protein appear to be minimal.	

Table S19 Bioinformatics Summary Results of Protein Band #8. The 248 AA Triose Phosphate isomerase protein, GI: 238032989 was compared with AOL v16 and with NCBI Protein on 17th May, 2016.

AOL full FASTA3	Four high-scoring alignments were identified with <i>E</i> scores smaller than 9.3e-51 and 50-53% identities over nearly the full length of four triosephosphate-isomerase proteins of diverse sources (wheat Tri a 31, two isoforms of Der f 25 from house dust mite, and shrimp Cra c 8). Yet triosephosphate isomerases are ubiquitous and there is no published evidence for such wide-spread cross-reactivity of this enzyme. Thus, these alignments are unlikely to represent a risk of possible cross-reactivity.
AOL 80mer FASTA3	The same four proteins were identified. The highest alignment was 62% identity.
AOL Exact 8 AA	A number of exact 8 AA matches were identified with the same four proteins.
NCBI BLASTP Allergen	Similar identity matches (51-52% over full-lengths) were found by BLASTP to triosephosphate isomerase of house dust mites and scabies. However, there is no evidence these proteins are cross-reactive with the <i>Pichia</i> protein.
NCBI BLASTP Toxin or Toxic	Moderately high scoring alignments were identified using toxin and toxic keywords. The matches with <i>E</i> scores of 2e-55 to 7e-65 and identities of 40% to 50% were found to homologous proteins from bacteria (e.g. <i>Escherichia coli</i> , <i>Bordetella sp.</i> , <i>Clostridium sp.</i>). No published evidence was found that these proteins are toxins.
NCBI BLASTP No keyword	Many matches of very high scoring identities with homologous proteins from other yeasts and molds were identified. The highest scoring alignment with a <i>Saccharomyces sp.</i> was with an <i>E</i> score of 1e-128 and identity of 71% over 248 AA.
Conclusion: It is unlikely that there is any risk of allergy for this protein. The same is true for toxicity. In addition, there are very good alignments to homologous proteins from many molds including <i>Saccharomyces sp.</i> Therefore, risks from consumption of this protein appear to be minimal.	

Table S20 Bioinformatics Summary Results of Protein Band #9a. The 161 AA hypothetical protein, GI:328350030 (cyclophilin) was compared with AOL v16 and with NCBI Protein on 17th May, 2016.

AOL full FASTA3	Seven relatively high-scoring alignments were identified with <i>E</i> scores smaller than 1e-46 to cyclophilin proteins of different sources (Fungus Asp f 27, House dust mite Der f 29, Yeast Mala s 6, unassigned allergen from plant <i>Catharanthus</i> , Birch Bet v 7, unassigned allergen from carrot <i>Daucus</i> , and Fungus Asp f 11. Yet cyclophilins are ubiquitous and while there is evidence of <i>in vitro</i> cross-reactivity, there is no published evidence for such wide-spread clinical cross-reactivity. Thus, these alignments are unlikely to represent a risk of possible clinical cross-reactivity.
AOL 80mer FASTA3	The same seven proteins were identified as having alignments with >35% identity.
AOL Exact 8 AA	All of the same seven proteins had 8 AA matches.
NCBI BLASTP Allergen	A few cyclophilins were identified with high matches (71% identity) of diverse allergenic sources (e.g. mold, <i>Aspergillus fumigatus</i> and <i>Malazessia furfur</i> ; mites, <i>Suidasia medanensis</i> and <i>Dermatophagoides farina</i> ; liver fluke, <i>Clonorchis sinensis</i> . However, there is not clear published proof that these proteins are clinically cross-reactive from such diverse sources.
NCBI BLASTP Toxin or Toxic	Modest scoring alignments were identified using toxin and toxic keywords. The matches with <i>E</i> scores of 3e-32 or larger and identity scores of approximately 40% to 50% identities to homologous proteins from bacteria (e.g. <i>Enterococcus sp.</i> , <i>Legionella sp.</i> , <i>Corynebacterium sp.</i> , also with some unusual species in terms of toxicity, such as <i>Danio rerio</i>). No published evidence was found that these proteins are toxins.
NCBI BLASTP No keyword	Many matches of very high scoring identities with homologous proteins from other yeasts and molds were identified. The best alignment was from <i>Saccharomyces sp.</i> with <i>E</i> score of 9e-88 and identity of 74% over
Conclusion: Cyclophilins are unlikely to pose a risk of allergy, although there can be some shared IgE binding. Toxicity is unlikely, as well since there are very good alignments to homologous proteins from many molds including <i>Saccharomyces sp.</i> Therefore, risks from consumption of this protein appear to be minimal.	

Table S21 Bioinformatics Summary Results of Protein Band #9b. The 154 AA cytosolic superoxide dismutase protein, GI: 238034030 was compared with AOL v16 and with NCBI Protein on 17th May, 2016.

AOL full FASTA3	Twenty three isoforms of olive pollen superoxide dismutase Ole e 5 were aligned with nearly identical scores relative to this <i>P. pastoris</i> protein #9b. Ole e 5 was first characterized by immunoblotting in 1998 by Boluda [12], and Butteroni demonstrated IgE binding using sera from a large number of olive pollen allergic subjects in 2005 [13], but till now this allergen has only been characterized by IgE tests without publication demonstrating biological activity [14].
AOL 80mer FASTA3	The same 23 were identified as having alignments with >35%. The highest alignment was 87% and all had 80 identities greater than 78%.
AOL Exact 8 AA	All of the same 23 proteins had 8 AA matches.
NCBI BLASTP Allergen	Multiple isoforms of olive pollen (Ole e 5) were identified with <i>E</i> scores ranging from 3e-55 to 9e-55 and approximately 57% identity over 152 AA. As noted above, there is evidence of IgE binding to the olive pollen using olive-pollen allergic subjects, but no demonstration of direct biological reactivity.
NCBI BLASTP Toxin or Toxic	The matches were identified to homologous proteins from bacteria (e.g. <i>Clostridium sp.</i> , <i>Corynebacterium sp.</i>), with high identities and significant <i>E</i> scores. However, the linkage appeared to be due the ability of the homologous proteins in a wide array of organisms to de-toxify free radicals. No published evidence was found that these proteins are toxins.
NCBI BLASTP No keyword	High scoring identity matches with homologous proteins from other yeasts and molds were identified. The highest scoring alignment with a <i>Saccharomyces sp.</i> with 79% identity over 154 AA.
Conclusion: The protein #9b, superoxide dismutase of <i>Pichia pastoris</i> is unlikely to pose a risk of allergy. The same is true for toxicity. In addition, there are very good alignments to homologous proteins from many molds including <i>Saccharomyces sp.</i> Therefore, risks from consumption of this protein appear to be minimal.	

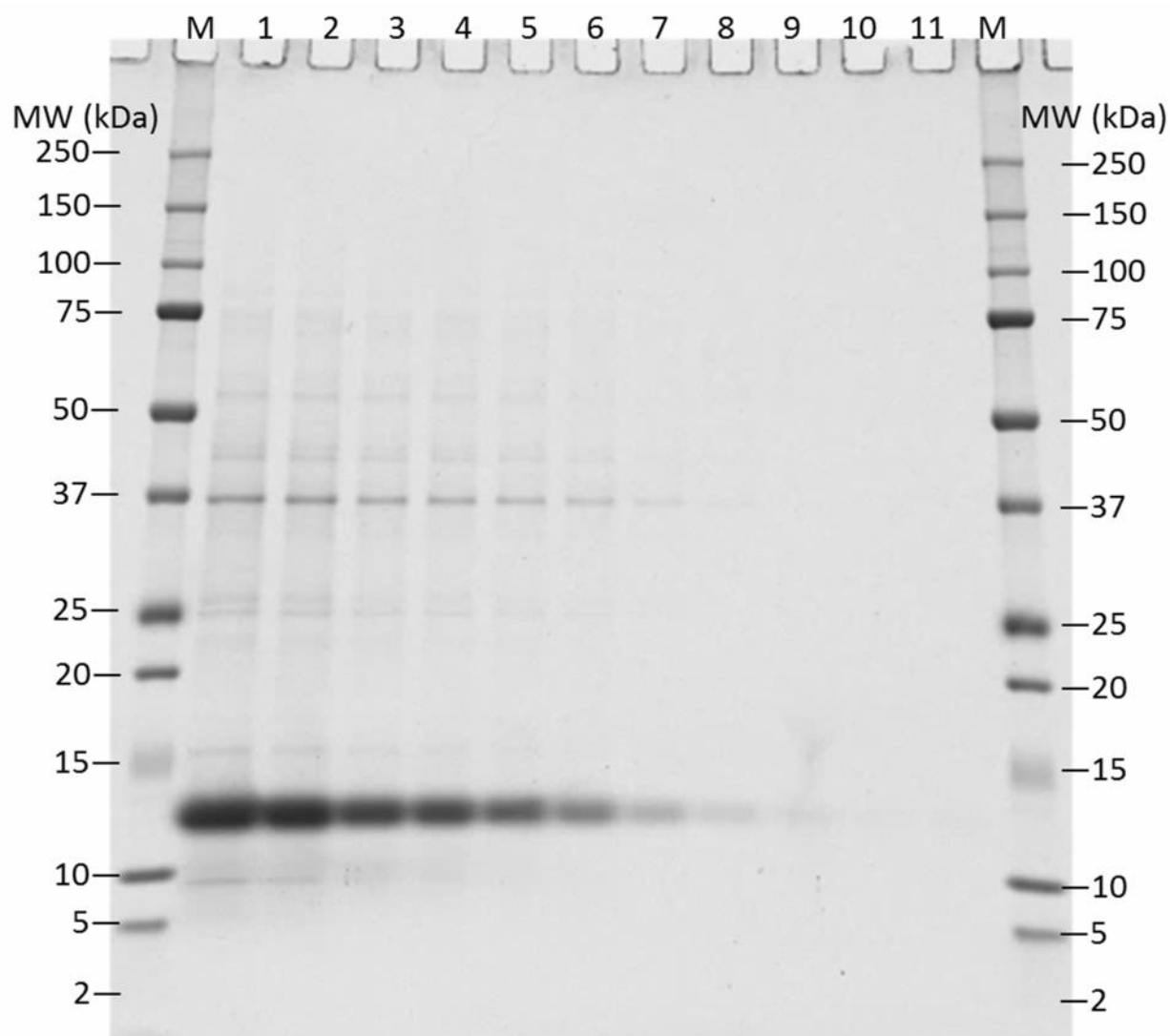
Table S22 Bioinformatics Summary Results of Protein Band #10. The 84 AA hypothetical protein

(mitochondrial ATPase inhibitor) was compared with AOL v16 and with NCBI Protein on 17th May,

2016.

AOL full FAST A3	Relatively low-scoring alignments were identified with <i>E</i> scores between 0.1 and 10 and identities of 19-30% that were primarily identified to the muscle protein, tropomyosin. The alignments are clearly poor and do not represent conservation of homology. Thus, it is unlikely to represent a risk of cross-reactivity.
AOL 80mer FASTA3	No matches were identified with >35% identity in any 80 AA match.
AOL Exact 8 AA	No exact 8 AA matches were identified.
NCBI BLASTP Allergen	No significant alignments were identified.
NCBI BLASTP Toxin or Toxic	No significant alignments were identified.
NCBI BLASTP No keyword	A few significant alignments were found for yeasts and molds. The highest scoring alignment with a <i>Saccharomyces sp.</i> was with an <i>E</i> score of 3e-20 and identity of 62% over 64 AA.
Conclusion: The ATPase inhibitors do not appear to be allergens or toxins. The highest identity score to <i>Saccharomyces sp.</i> was with an <i>E</i> score of 3e-20 and 49% identity over 69 AA. Thus, it is very unlikely that this protein is an allergen or a toxin. The risks from consumption of this protein appear to be minimal.	

Figure S1 Coomassie Brilliant Blue Stained SDS-PAGE Gel Showing the Serial Dilution of *P. pastoris* Produced LegHb Starting from 200% to 2.5% of Total Protein. Lane M, molecular weight marker; Lane 1, 200% of total protein (2.96 μg); Lane 2, 150% of total protein (2.21 μg); Lane 3, 100% of total protein (1.47 μg); Lane 4, 80% of total protein (1.18 μg); Lane 5, 60% of total protein (0.88 μg); Lane 6, 40% of total protein (0.59 μg); Lane 7, 20% of total protein (0.29 μg); Lane 8, 10% of total protein (0.15 μg); Lane 9, 5% of total protein (0.075 μg); Lane 10-11, 2.5% of total protein (0.037 μg).



Supporting references

- [1] Banerjee, H. n., Verma, M., Search for a novel killer toxin in yeast *Pichia pastoris*. *Plasmid* 2000, 43, 181-183.
- [2] Chessa, R., Landolfo, S., Ciani, M., Budroni, M., *et al.*, Biotechnological exploitation of *Tetrapisispora phaffii* killer toxin: heterologous production in *Komagataella phaffii* (*Pichia pastoris*). *Applied Microbiology and Biotechnology* 2017, 101, 2931-2942.
- [3] Comitini, F., Mannazzu, I., Ciani, M., *Tetrapisispora phaffii* killer toxin is a highly specific beta-glucanase that disrupts the integrity of the yeast cell wall. *Microbial cell factories* 2009, 8, 55.
- [4] Assarehzadegan, M. A., Sankian, M., Jabbari, F., Tehrani, M., *et al.*, Identification of methionine synthase (Sal k 3), as a novel allergen of *Salsola kali* pollen. *Molecular Biology Reports* 2011, 38, 65-73.
- [5] Lauble, H., Kennedy, M., Emptage, M., Beinert, H., Stout, C., The reaction of fluorocitrate with aconitase and the crystal structure of the enzyme-inhibitor complex. *Proceedings of the National Academy of Sciences* 1996, 93, 13699-13703.
- [6] Chiu, L. L., Lee, K. L., Lin, Y. F., Chu, C. Y., *et al.*, Secretome analysis of novel IgE-binding proteins from *Penicillium citrinum*. *PROTEOMICS-Clinical Applications* 2008, 2, 33-45.
- [7] Galli, B. D., Martin, J. G. P., da Silva, P. P. M., Porto, E., Spoto, M. H. F., Sensory quality of Camembert-type cheese: Relationship between starter cultures and ripening molds. *International Journal of Food Microbiology* 2016, 234, 71-75.
- [8] Kitamoto, K., Cell biology of the Koji mold *Aspergillus oryzae*. *Bioscience, Biotechnology, and Biochemistry* 2015, 79, 863-869.
- [9] Shen, H., Choo, K., Tang, R., Lee, C., *et al.*, Allergenic components of *Candida albicans* identified by immunoblot analysis. *Clinical & Experimental Allergy* 1989, 19, 191-196.
- [10] Shen, H.-D., Choo, K.-B., Lee, H.-H., Hsieh, J. C., *et al.*, The 40-kilodalton allergen of *Candida albicans* is an alcohol dehydrogenase: molecular cloning and immunological analysis using monoclonal antibodies. *Clinical & Experimental Allergy* 1991, 21, 675-681.
- [11] Onishi, Y., Kuroda, M., Yasueda, H., Saito, A., *et al.*, Two-dimensional electrophoresis of *Malassezia* allergens for atopic dermatitis and isolation of Mal f 4 homologs with mitochondrial malate dehydrogenase. *European Journal of Biochemistry* 1999, 261, 148-154.
- [12] Boluda, L., Alonso, C., Fernández-Caldas, E., Purification, characterization, and partial sequencing of two new allergens of *Olea europaea*. *Journal of Allergy and Clinical Immunology* 1998, 101, 210-216.
- [13] Butteroni, C., Afferni, C., Barletta, B., Iacovacci, P., *et al.*, Cloning and expression of the *Olea europaea* allergen Ole e 5, the pollen Cu/Zn superoxide dismutase. *International Archives of Allergy and Immunology* 2005, 137, 9-17.
- [14] Esteve, C., Montealegre, C., Marina, M., García, M., Analysis of olive allergens. *Talanta* 2012, 92, 1-14.