Engineering the Common Soybean (*Glycine max*) to Biosynthesize Alpha-lactalbumin

Elena Statham

Department of Molecular, Cellular, and Developmental Biology

University of Colorado Boulder

April 2nd 2024

Defense Committee:

Thesis Advisor: Brian DeDecker, Department of Molecular, Cellular, and Developmental Biology

Honors Council Representative: Christy Fillman, Department of Molecular, Cellular, and Developmental Biology

Outside Reader: Andrew Martin, Department of Ecology and Evolutionary Biology

Abstract

6 billion people drink cow's milk worldwide. There are tons of dietary benefits to cow's milk as it is a strong source of proteins, fats, carbohydrates, and nutrients, and it contains all the essential amino acids. The dairy industry is booming, producing hundreds of million metric tons of milk worldwide every year. The cows producing this milk often undergo abuse such as continuous forced impregnation, over-milking, malnourishment, and restrictive confinement. Additionally, the dairy farm industry is responsible for approximately 2.9% of all human-induced greenhouse gas emissions due to the methane cows produce. To reduce the abuse of the dairy industry and reduce greenhouse gas emissions other ways of producing cow's milk are being researched. Current research has been aimed at synthetically producing the major proteins, particularly the whey proteins, of cow's milk in alternative host organisms such as yeast and *E. coli*. However, these organisms are unable to produce such proteins at a cost-effective and large enough scale to be profitable over current dairy cow farming. Little research has been done outside of these microbial hosts. We hypothesize that *Glycine max,* the common soybean, can efficiently produce the whey proteins of cow's milk, with cheaper production and larger yields than current methods. This research found that the whey protein, α-lactalbumin, can be transformed into agrobacterium, but further experiments are needed to determine α-lactalbumin protein production inside a soybean. The costs and yields available through this synthetic farming method would then be researched for possible industrial production of synthetic cow's milk.

Introduction

Approximately 6 billion people worldwide drink cow's milk (Visioli & Strata, 2014). Milk is a dietary staple around the world and provides important nutrients to the human diet, as it contains proteins, fats, carbohydrates, and nutrients important for bone health (*Milk - Better Health Channel*, n.d.). Compared to plant-based milks dairy milk contains more energy, fat, carbohydrates, vitamins C, $B_{2,}$ $B_{12,\text{ and}}$ A, biotin, pantothenic acid, calcium, phosphorus, and iodine (Walther et al., 2022). Dairy products have been associated with many health benefits when consumed regularly, such as a possible protective benefit against stroke, reduced risk of colon cancer, association with lower blood pressure, protection against type 2 diabetes, and protection against tooth decay. When dairy products are removed from the diet it can lead to calcium deficiency, which can become a serious health concern for women over the age of 50 and the elderly as it can lead to Osteoporosis (*Milk - Better Health Channel*, n.d.). While plant-based milks contain many nutrients from dairy milk, they often lack some key nutrients and contain nutrients at lower concentrations than dairy milk. Dairy milk is an important source of nutrition in the human diet and helps keep billions of people healthy.

To supply this milk, dairy farms produced 544 million metric tons of cow's milk worldwide in 2022 (*U.S. Dairy: Milk Produced per Cow 2023 | Statista*, n.d.). The average cow emits between 154 to 264 pounds of methane each year (*Agriculture and Aquaculture: Food for Thought | US EPA*, n.d.). The environmental impacts of dairy farming are responsible for 2.9% of total human-induced greenhouse gas emissions (*What Are the Environmental Impacts of Dairy Farming?*, 2022). Not only is the dairy industry contributing to greenhouse gas emissions, but it also subjects cows to cruel treatment. Dairy cows are repeatedly impregnated, forced to overproduce milk, malnourished, and confined to tight living spaces (*Cows | The Humane*

Society of the United States, n.d.). Cows are not treated as living animals but as vectors for mass profit, all the while this industry contributes to the climate crisis.

Not producing milk is not a viable solution to this climate and animal cruelty issue because dairy milk is key to many individuals' nutritional needs. Instead of reducing milk production, milk could be synthesized through other organisms, cutting out cows from the equation. Many companies are currently producing the whey proteins found in cow's milk. Whey proteins account for 20% of all proteins in cow's milk. The two main whey proteins are α-lactalbumin and β-lactoglobulin. The synthesis of whey is an important starting point for synthetic cow milk because whey proteins play an important role in the human diet as these proteins contain the 9 essential amino acids, functional peptides, antioxidants, and immunoglobins (Patel, 2015). Companies currently synthesizing whey are doing so by using yeast and *E. Coli* systems. However, yeast has disadvantages such as containing active proteases that degrade foreign proteins, they accept fewer vectors, and their mechanisms of protein expression are not completely understood (*Yeast Expression - Creative Biostructure*, n.d.). *E. coli* used for synthetic protein engineering is unable to catalyze disulfide bond formation, it's unable to glycosylate, and it is difficult to get a high protein expression (Bhatwa et al., 2021).

A much more efficient synthetic protein host is *glycine max,* the common soybean. A major concern for most GMO crops is cross-contamination, but soybeans are less likely to cross-pollinate because they have heavy pollen granules that do not travel far. Approximately 5 meters of distance is required between GMO soybean fields and non-GMO soybean fields. Comparatively, other GMO crops need to be separated from other crops by 10m-150m (Clay, 2019). Soybeans are nitrogen-fixing organisms. This means they naturally fertilize the soil, making it efficient to crop rotate without the need to fertilize soil between growing seasons.

Additionally, soybeans do not need extra material to grow. Unlike yeast and *E. coli* which need to be supplied with nutrients to grow, soybeans only need water and sunlight, which is readily available in industrial-scale agricultural areas, such as Iowa. Soybeans are part of industrial farming in the U.S., meaning the infrastructure is already set up for large yields.

Glycine max is hypothesized to be a suitable host for whey protein synthesis. To test this hypothesis one of the two main whey proteins, α-lactalbumin, was chosen for protein synthesis. α-lactalbumin is made of 123 amino acids. 45% of the protein structure is alpha helices and 10% are beta-sheets. Its domains are connected by the disulfide bonds (*Alpha_Lactalbumin*, n.d.).

To test this hypothesis plasmids were designed in *E. coli* and sequence verified. The final plasmid was transferred to *A. tumefaciens* (agrobacterium) through transformation. *A. tumefacient* have the natural ability to insert their DNA into the soybean, leaving the soybean with the engineered α-lactalbumin. The genetically modified soybean now has α-lactalbumin DNA to transcribe and translate into a functional protein. Once the plant has grown, tissues of the plant will be extracted and tested for α-lactalbumin protein presence using a Western Blot.

Whey protein from cows milk synthesized in soybeans

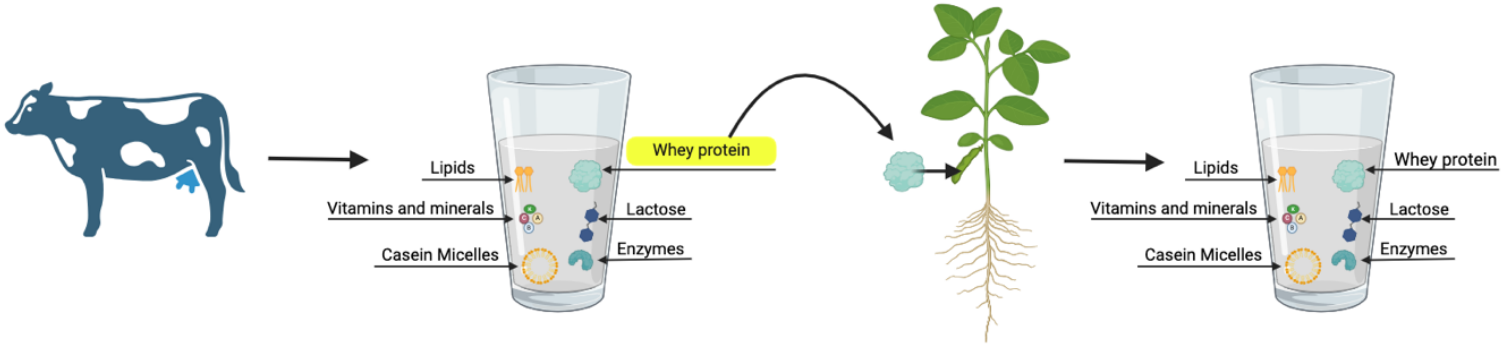

Figure 1: The objective of this research is to take the α-lactalbumin whey protein found in cow's milk and produce it inside a soybean plant. The long-term result being chemically identically dairy milk farmed from soybeans.

Materials and Methods

Plasmid Design

The α-lactalbumin gene was designed to translate into the genome of soybeans. The α-lactalbumin gene itself was sourced from Uniprot.org (Accession ID P00711) in Bovine (*LALBA - Alpha-Lactalbumin - Bos Taurus (Bovine) | UniProtKB | UniProt*, n.d.). The entire gene was codon optimized for *Glycine max*. The signal sequence of the gene that localizes where proteins will end up in the host organism was switched from the original bovine signal sequence to the soybean signal sequence (amino acids: MKVAFAAVLLICLVLSSSLFEVSMA). The bovine signal sequence allows the α-lactalbumin proteins to export out of the mammalian cells and into the mammary glands of cows. This signal sequence was cut off and replaced with a plant signal sequence called explant which is hypothesized to export α-lactalbumin to the apoplast of the soybean plant cells. A TEV protease cleavage site, found on neb.com, was added to the end of the gene. This TEV sequence has been shown to increase protein stability and it contains a

Histidine tag that will be used as an antibody binding site in a Western Blot (*TEV Protease | NEB*, n.d.). The α-lactalbumin gene construct was ordered as a linear DNA fragment that contained Type II restriction enzyme cut sites from Twist Biosciences. The final plasmid was constructed with a RUBY marker that turns the α-lactalbumin protein red in the mature plant (He et al., 2020).

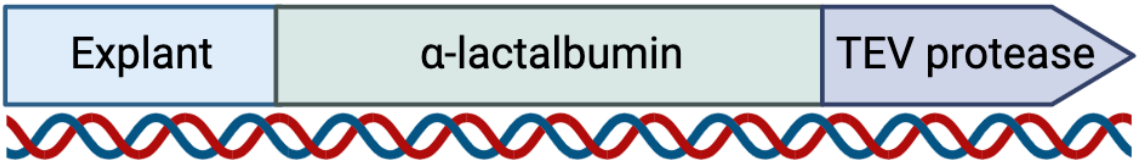

Figure 2: Gene ordered from Twist Biosciences constructed with an apoplast localization tag, α-lactalbumin protein coding sequence, and a TEV protease containing a His tag.

Plasmid Construction

A total of three intermediate plasmids of various antibiotic resistance were used in combination to assemble three separate gene fragments into a single DNA plasmid. The first plasmid was an ampicillin-resistant plasmid in which the ordered α-lactalbumin gene construct was inserted. The plasmid was sequence verified using the services of Plasmidsaurus whole plasmid sequencing as well as sequence alignment on Benchling. The α-lactalbumin gene was then transferred to the second intermediate plasmid that consisted of a promoter (Gmubi) and terminator (HSPt) in a chloramphenicol-resistant backbone. Next, the promotor-α-lactalbumin-terminator sequence was inserted alongside the RUBY marker sequence into its final plasmids. The final plasmid, pBEHA2, was a kanamycin-resistant plasmid suitable for agrobacterium transfection. All backbones contained a blue/white gene that was removed when restriction enzymes correctly cut the plasmid and the intended gene insert was ligated into the backbone.

This blue/white marker allows colonies with the intended insert to appear white on LB agar plates.

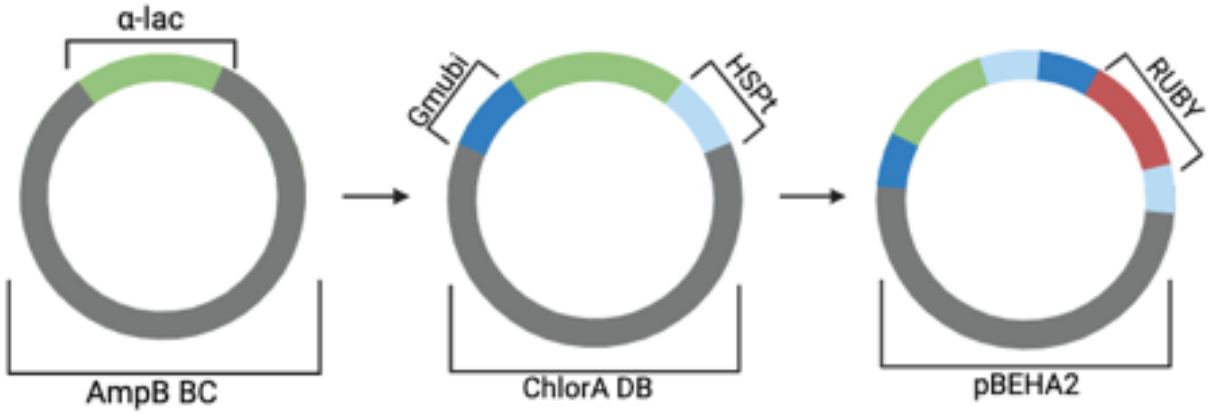

Figure 3: The design of all three plasmids. Gmubi is the promotor and HSPt is the terminator for α-lactalbumin and RUBY. All backbones are compatible in *E. coli* and pBEHA2 is also compatible to grow in agrobacterium.

Golden Gate Cloning

Each reaction was a total of 20uL in volume, consisting of 0. 05pmol backbone and 0.1pmol of intended DNA insert, 0.75uL type II restriction enzyme (BsaI, PaqC1), 0.75uL enzyme activator (for PaqC1), 1uL T4 ligase, 2uL10x ligase buffer, and sterile ddH2O until 20uL is reached. All reactions were incubated at 37°C for 1-18hrs and then stored at 4°C.

***E. coli* Transformation and Growth Conditions**

Following the golden gate reaction, 5uL of NEB chemically competent *E. coli* cells are thawed on ice for 10 minutes. Once thawed 0.5uL of DNA from the golden gate reaction is added to the cells and placed on ice for 30 minutes then heat shocked at 42°C for 30 seconds and

95uL of LB broth is added to the *E. coli*/DNA solution. If in the ampicillin-resistant backbone, the solution can be plated immediately, when in a chloramphenicol or kanamycin-resistant backbones the solution was incubated at 37°C for one hour before being plated. The plates consisted of LB agar, XGAL, IPTG, and corresponding antibiotics. After spreading media on plates, plates were incubated at 37°C overnight. After colonies appeared on the plates, cell grow-ups were made by selecting a white colony with a sterile pipette tip and dropping it into 5mL of LB broth combined with 5uL of appropriate antibiotic. These cell grow-ups were then placed in 37°C incubators overnight on an orbital shaker at 260RPM.

Plasmid Isolation and Quantification

DNA plasmids were isolated from the overnight cell grow-ups following the Qiagen miniprep plasmid isolation kit protocol. The alterations made to this protocol were the PB buffer step being excluded and DNA was eluted with 50uL ddH_2O instead of EB buffer. A nanodrop instrument was used to identify the DNA concentration of each sample.

Restriction Enzyme Digest and Gel Electrophoresis

Plasmids were analyzed for accuracy through restriction digests and gel electrophoresis. Digests consist of 20uL reaction in a PCR tube that contains 2uL 10x cutsmart buffer, 3uL miniprep DNA, 0.5uL restriction enzyme (Bsa, PaqC1, BaeI, PstI), then incubated at 37°C for one hour. The samples ran on a 5% agarose gel with a Life 1kb ladder. The size of the DNA bands in the gel represents if the DNA is inserted correctly into the backbone. The DNA was

then sent off to Plasmidausaurus and sequenced. The sequence was aligned on Benchling to confirm the complete accuracy of the plasmids.

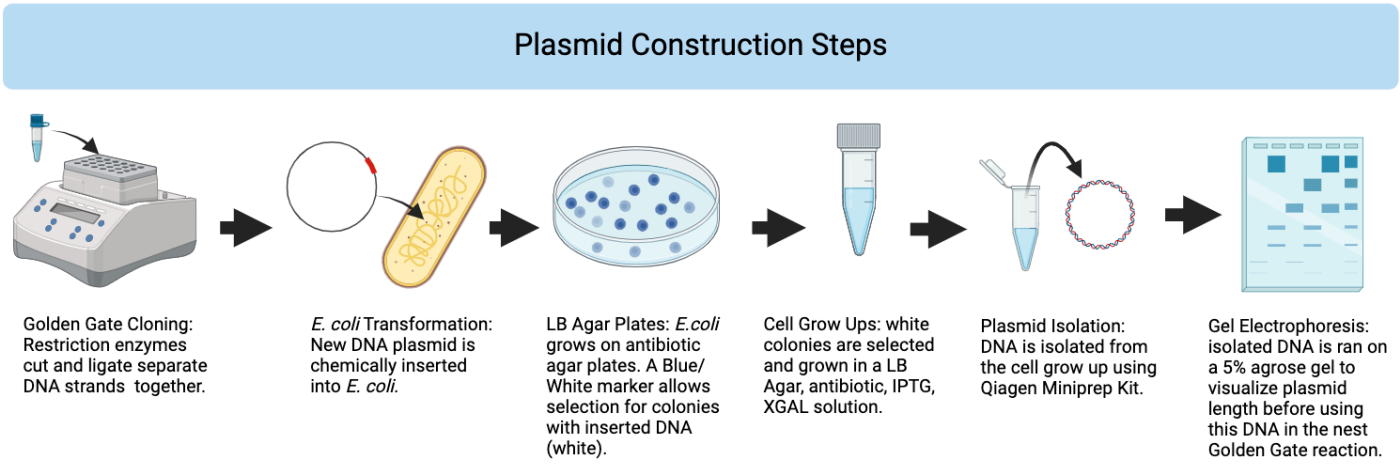

Figure 4: Plasmid construction in *E. coli* is performed in 6 experiments. First, a golden gate reaction cuts and ligates DNA into one plasmid. Second, a transformation inserts the new plasmid into *E. coli* cells. Third, *E. coli* cells grow on agar plates and white colonies represent cells containing the new plasmid. Fourth, a white colony is grown in LB/antibiotic solution. Fifth, DNA is isolated from the cells. Sixth, DNA is ran on a gel electrophoresis for analysis of plasmid length.

Agrobacterium Mediated Transformations

The final plasmid in the pBEHA2 backbone was electroporated into EHA 105 agrobacterium stock at 50uL. After electroporation, the solution was resuspended in 200uL of LB medium and plated onto LB agarose plates that contained kanamycin. Plates were incubated at 27°C for two days and then colonies were selected for overnight grow-ups in a solution of 5mL LB broth and 5μL kanamycin. The DNA from the cell culture grow-ups was isolated using the Qiagen miniprep kit. The isolated DNA was then transformed back into *E. coli*, plated, grown up, and again DNA isolated. This DNA was then sent to Plasmidausaurus to be sequence verified.

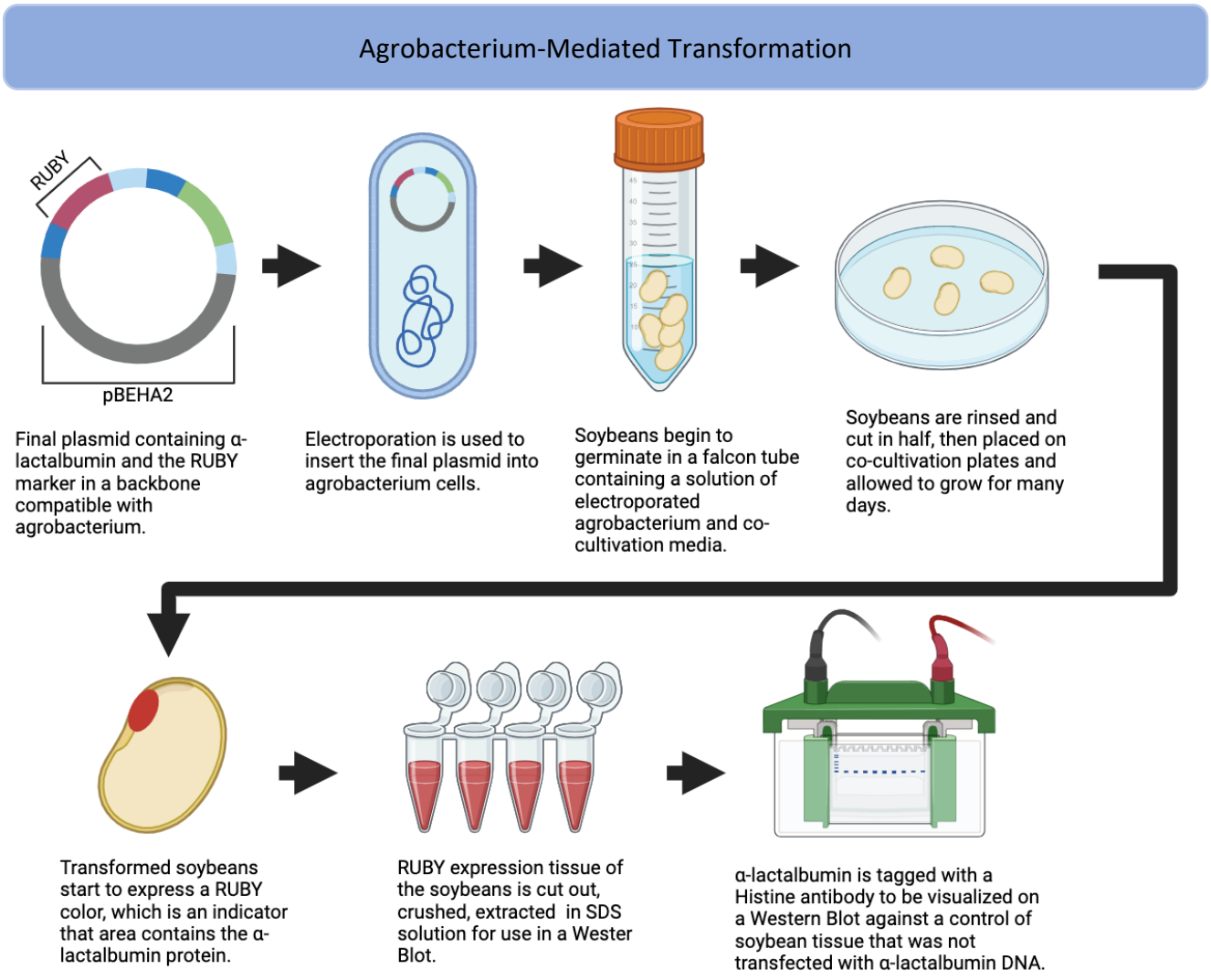

Figure 5: The final α-lactalbumin plasmid is inserted into agrobacterium cells using electroporation. The agrobacterium containing α-lactalbumin is then put into solution with co-cultivation medium and embryonic soybeans. The soybeans are later washed, cut in half, and placed cut side down on plates to germinate. Growing beans if transfected will begin to produce α-lactalbumin protein which appears as red tissue on the soybeans. This red tissue is cut out and digested in an SDS solution to be ran on a Western Blot that tags the α-lactalbumin protein with a Histidine antibody. Further verification of the protein could be done through mass spectrometry analysis to confirm results.

Results

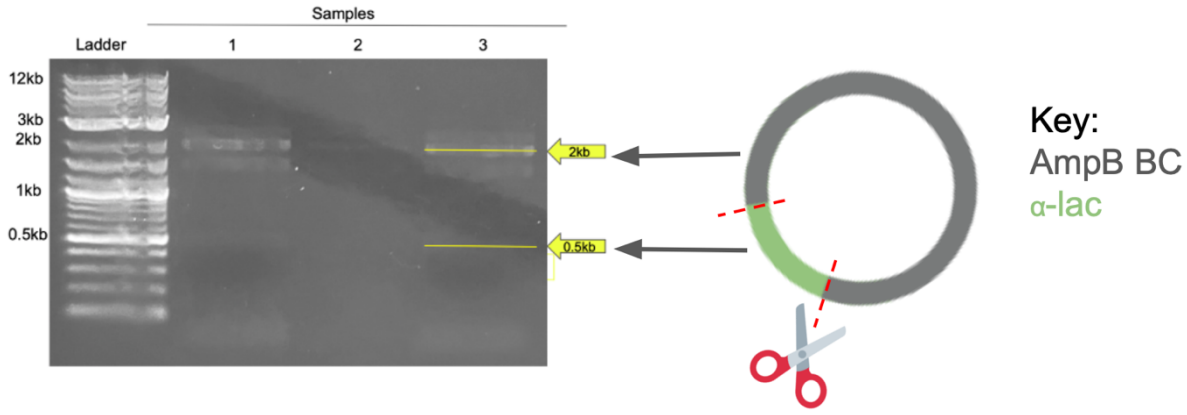

Figure 6: 5% agarose gel used to run plasmid AmpB BC to visualize possible α-lactalbumin integration. Samples 1-3 are all different colonies of the plasmid after digestion with PaqC1.

The α-lactalbumin gene was integrated into the ampicillin-resistant backbone AmpB BC. Figure 6 shows the 5% agarose gel which ran three separate samples of the reaction after being digested by PaqC1. As seen in the gel, two bands were visualized. The band at 2kb is representative of the entire backbone excluding the α-lactalbumin gene. The second band at 0.5kb is representative of only the α-lactalbumin gene. These three samples were sent off for sequencing and the sequencing results were aligned to the expected DNA on Benchling. The alignment showed that sample three aligned perfectly.

The verified α-lactalbumin in AmpB BC was then used in the next golden gate reaction, so that α-lactalbumin would be integrated into a chloramphenicol backbone, ChlorA DB, with a promoter and terminator. Proceeding the golden gate, a transformation was done and plated, white colonies were selected for an overnight grow-up, then the DNA was isolated, digested, and ran on a gel (Figure 7).

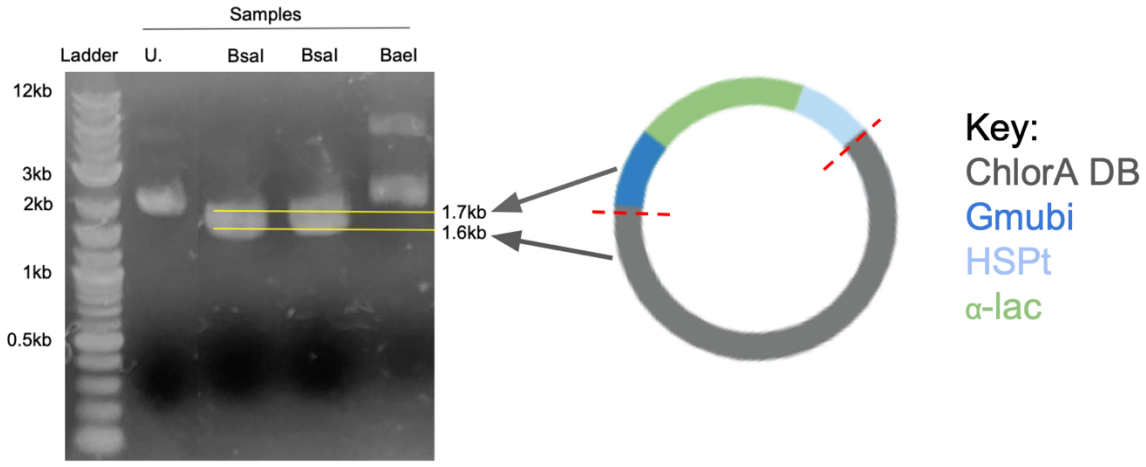

Figure 7: 5% agarose gel ran plasmids ChlorA DB with promoter, a α-lactalbumin, and terminator sequences. Sample 1 is the plasmid before digestion by a restriction enzyme, lane 2. Both BsaI samples are the plasmids digested with restriction enzyme BsaI and the last lane is the plasmid digested with restriction enzyme BaeI.

In Figure 7, the large band on the agarose gel in lanes 3 and 4 was suggestive of two bands that appear as one because the restriction enzyme BsaI cut the plasmid almost exactly in half (one piece at 1.7kb and the other at 1.6kb). The BaeI cut sites, lane 5, did not reflect the expected DNA length when run on the gel. This was likely due to background DNA contamination, as the expected length was much larger than the uncut plasmid. Because the BsaI lanes only show one large band this suggested that α-lactalbumin successfully integrated into the Chlor backbone.

The final plasmid assembly takes the promoter, α-lactalbumin, and terminator sequence out of ChlorA DB and into the final backbone pBEHA2. Along with α-lactalbumin, the protein marker RUBY is added to the golden gate reaction so that both RUBY and α-lactalbumin will be in one final pBEHA2 plasmid. Once the final plasmid construction went through golden gate

assembly, transformation, plating, cell grow-ups, and DNA isolation, the digested DNA ran on an agarose gel (Figure 8).

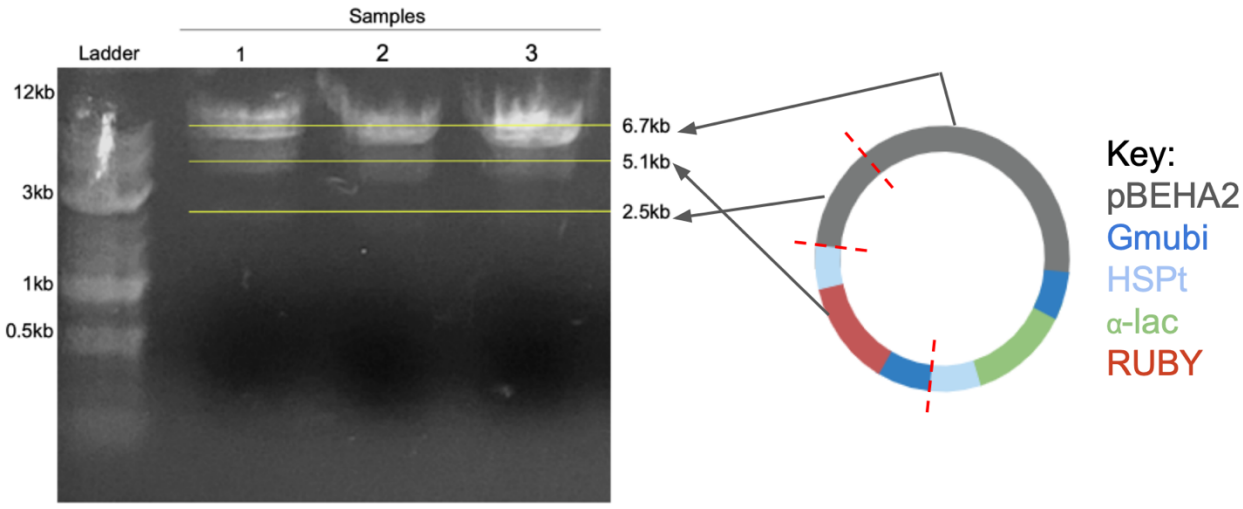

Figure 8: 5% agarose gel that ran a pBEHA2 plasmid containing promoters, terminators, α-lactalbumin, and RUBY marker. All three samples were digested with restriction enzyme PstI.

As seen in Figure 8 above, all three samples were cut 3 times by PstI and resulted in the expected DNA lengths of 6.7bp, 5.2bp, and 2.5bp. Samples 1 and 2 were both sequenced by Plasmidsaurus and aligned perfectly on Benchling. All the DNA needed for successful bean transfections was now in one plasmid. The next step was to insert this plasmid into the agrobacterium, which was done through electroporation into the agrobacterium cell line *A. tumefaciens* EHA 105. The electroporated EHA 105 cells were plated and grown up. Part of the cell grow-up was made into glycerol stocks to ensure the cell line was stable and could be used later. The rest of the cell line went through a mini-prep to isolate the DNA. To ensure the agrobacterium did not mutate, the DNA was transformed back into *E. coli*. After transforming,

plating, grow-ups, and isolating the DNA, the DNA was sent to Plasmidsauras for sequencing.

Figure 9: Benchling alignment of the expected final pBEHA2 plasmid sequence (top) and the sequence returned by Plasmidauras (bottom).

As seen in Figure 9 above, the pBEHA2 plasmid was correctly integrated into agrobacterium with no mutations. The next step is to transform soybeans with the agrobacterium containing the synthetic plasmid. These beans will grow up and present a red/ruby color where the α-lactalbumin protein is present. The ruby tissue will be cut off the plant and a western blot will run the transformed plant samples next to a plant transformed without the α-lactalbumin sequence. The western blot would confirm if α-lac protein is successfully produced by these GMO soybeans. For the next steps, see Figure 5.

Discussion

The final pBEHA2 plasmid containing the α-lactalbumin sequence and RUBY marker sequence with promoters and terminators was successfully transformed into agrobacterium and verified through Plasmidsauras sequencing and a Benchling alignment.

Dairy milk provides many nutrients and proteins at higher concentrations than plant-based milks and has many health-associated benefits when consumed regularly. Whey proteins are a predominant protein found in cow's milk and can also be produced in *E. coli* and yeast. Current whey protein manufacturing takes lots of resources and money through such hosts. Alternatively, the production of whey proteins in soybeans is a promising mechanism for producing synthetic whey and eventually synthetic cow's milk. This is because soybeans are part of a large-scale farming infrastructure, they take greenhouse gasses out of the atmosphere, and they are nitrogen-fixing organisms. Soybeans as a host for dairy milk production have the potential to be a large-scale, low-cost, environmentally sustainable, and humane animal welfare alternative to current dairy farming practices.

Further research needs to be completed to fully integrate the α-lactalbumin plasmid from agrobacterium into soybean seedlings (see Figure 5). Once verification through a Western Blot proves successful α-lactalbumin protein production, then an assessment of production yield efficiency can be completed. To increase whey yields in soybeans, research may investigate the alteration of amino acid producing cellular mechanisms. If the soybean can be altered to produce more of the amino acids used in α-lactalbumin then the cells may have the ability to increase α-lactalbumin protein production as well.

These results in combination with future experimentation show promise that a genetically transformed soybean may generate the whey protein α-lactalbumin. This is the first step of many to the production of a chemically identical form of dairy milk that is synthesized through soybeans. Dairy products synthesized from soybeans at lower monetary, environmental, and animal welfare costs could provide the nutrition of today's dairy milk to 6 billion people worldwide.

Citations

Agriculture and Aquaculture: Food for Thought | US EPA. (n.d.). Retrieved February 7, 2024, from https://www.epa.gov/snep/agriculture-and-aquaculture-food-thought

Alpha_Lactalbumin. (n.d.). Retrieved March 11, 2024, from https://collab.its.virginia.edu/access/content/group/f85bed6c-45d2-4b18-b868-6a2353586804/2/Ch06_Do_V_Alpha-Lactalbumin-_-/Ch06_Do_V_Alpha-Lactalbumin_Alpha_Lactalbumin.html#

Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X.-Z., & Zhou, T. (2021). Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. *Frontiers in Bioengineering and Biotechnology*, *9*, 630551. https://doi.org/10.3389/fbioe.2021.630551

Clay, D. E. (2019). *Chapter 23: Nitrogen Fixation*.

Davoodi, S. H., Shahbazi, R., Esmaeili, S., Sohrabvandi, S., Mortazavian, A., Jazayeri, S., & Taslimi, A. (2016). Health-Related Aspects of Milk Proteins. *Iranian Journal of Pharmaceutical Research : IJPR*, *15*(3), 573–591.

He, Y., Zhang, T., Sun, H., Zhan, H., & Zhao, Y. (2020). A reporter for noninvasively monitoring gene expression and plant transformation. *Horticulture Research*, *7*(1), Article 1. https://doi.org/10.1038/s41438-020-00390-1

LALBA - Alpha-lactalbumin—Bos taurus (Bovine) | UniProtKB | UniProt. (n.d.). Retrieved October 11, 2023, from https://www.uniprot.org/uniprotkb/P00711/entry

Milk—Better Health Channel. (n.d.). Retrieved March 5, 2024, from https://www.betterhealth.vic.gov.au/health/healthyliving/milk

Patel, S. (2015). Emerging trends in nutraceutical applications of whey protein and its derivatives. *Journal of Food Science and Technology*, *52*(11), 6847–6858. https://doi.org/10.1007/s13197-015-1894-0

TEV Protease | NEB. (n.d.). Retrieved February 29, 2024, from https://www.neb.com/en-us/products/p8112-tev-protease

U.S. dairy: Milk produced per cow 2023 | Statista. (n.d.). Retrieved February 7, 2024, from https://www.statista.com/statistics/194935/quantity-of-milk-produced-per-cow-in-the-us-since-1999/

Visioli, F., & Strata, A. (2014). Milk, Dairy Products, and Their Functional Effects in Humans: A Narrative Review of Recent Evidence1. *Advances in Nutrition*, *5*(2), 131–143. https://doi.org/10.3945/an.113.005025

Walther, B., Guggisberg, D., Badertscher, R., Egger, L., Portmann, R., Dubois, S., Haldimann, M., Kopf-Bolanz, K., Rhyn, P., Zoller, O., Veraguth, R., & Rezzi, S. (2022). Comparison of nutritional composition between plant-based drinks and cow's milk. *Frontiers in Nutrition*, *9*. https://doi.org/10.3389/fnut.2022.988707

Yeast Expression—Creative Biostructure. (n.d.). Retrieved March 6, 2024, from https://www.creative-biostructure.com/yeast-expression-522.htm