In Silico Cloning of a Phe-free κ-casein in *Kluyveromyces lactis* using cheese whey as c-source for the production of cream cheese for Phenylketonuria patients

Paola Sanabria Ramírez[1], Martha Diaz[2], Laura Buitrago[1], Maria Alejandra Escobar Holguin[1], Miguel Menco[2], and Javier Lopez[1]

[1]Affiliation not available
[2]University of Los Andes (Uniandes)

May 22, 2020

Abstract

Phenylketonuria (PKU) is a genetic disorder that increases the levels of Phenylalanine (Phe) in the blood due to a defect in the gene that helps create the enzyme necessary to break down this amino acid in the uptake of several protein products. In consequence, PKU patients have limited access to suitable protein sources to complement their daily diets. In this research, it is proposed the in silico cloning of a Phe-free k-casein as an alternative source of protein for PKU patients, and its use in a cream cheese formulation. The cloning process is proposed in *K. lactis* as hos microorganism using cheese whey as c-source and an expression vector built with from pKLAC2 integrated with a Phe-free κ-casein gene codon-optimized for its expression by secretion in *K. lactis*. Additionally, it is proposed a downstream separation process combining a centrifugation unit for the biomass separation, and filtration and ultrafiltration units for the Phe-free κ-casein purification. Finally, it was proposed a formulation for Phe-free cream cheese, including Phe-free k-casein (12% wt.), canola oil (14% wt.), lactase (3% wt.), salt (7% wt.) and water (csp).

Introduction

The recombinant κ-casein production has been studied since 1980. Initially, the κ-casein was cloned and expressed in *E. coli (Kang & Richardson, 1988; Oh & Richardson, 1991)*. In the last decade, the central part of research production has been focused on the expression of κ-casein in transgenic organisms as cattle, rabbits, and mice (Brophy et al., 2003; Gutiérrez-Adán et al., 1996).

Additionally, different authors have produced recombinant caseins for the treatment of Phenylketonuria (PKU) patients, which is a genetic alteration of the metabolism that affects 2 to 6:100,000 births (RevistaSemana, 2010). For example, ("Isolation and some effects of functional low-phenylalanine -casein expressed in the milk of transgenic rabbits", 2007) achieved at a high level of expression of a low-phenylalanine (Phe) $\varkappa$-casein in the milk of transgenic rabbits digestible with chymosin. Likewise, (Goda et al., 2000) produced of an intracellular recombinant Phe-free α_{s1}-casein in *Escherichia coli* using pCRII-TA as a cloning vector, and pET22b and pET23b as expressing vectors. However, the authors were not successful in expressing this protein in a eukaryotic system. Finally, (Appaiah & Vasu, 2020) cloned a genetically modified version of α-casein with increased content of large neutral amino acid (LNAA) in *E. coli* using pPICZ plasmid as vector and expressed in *Pichia pastoris*. The objective of include LNAA in the structure of α -casein was to blocks the passage of Phe into the brain PKU patients as a result of competition for the common transporter between Phe and LNAA in the gastrointestinal tract.

In general, the production of a Phe-free κ-casein could contribute significantly to obtain high-energy foods that play a unique role in nutrition and the prevention of significant neurocognitive deficits in PKU patients (Cleary & Skeath, 2019). Not to mention that the use of κ-casein as a protein source in new dairy products could expand the food options for patients with PKU, which today is limited (van & Ney, 2012).

On the other hand, Given its successful applications in the dairy industry, *Kluyveromyces lactis* has been considered a host for large-scale of recombinant proteins production by its high product yields, stable genomic integration of expression cassettes, cost-effectiveness, high cell density, protein maturation similar to mammalian cells, among others. Approximately 100 heterologous proteins produced by *K. lactis* have been reported (Spohner et al., 2016; *Recombinant Protein Production in Yeasts*, 2012).

K. lactis can grow on different substrates as cheese whey which is a by-product of the dairy industry with high protein and salts content that can easily use as a culture medium to develop products of high value-added (Ryan & Walsh, 2016; Yadav et al., 2015; Rubio-Texeira, 2006). Cheese whey represents 85–95 % of the milk volume and retains about 55 % of its nutrients, being its main constituents water (approximately 93 % of the total whey volume), lactose (70–72 % of the total solids), whey proteins (8–10 % of the total solids), and minerals (12–15 % of the total solids)(Ryan & Walsh, 2016; Tesfaye et al., 2019).

Nowadays, the lack of affordable methods for the use of cheese whey forces industries to dump most of this residue into wastewater and dry land, a fact that poses a permanent danger to the environment (Rubio-Texeira, 2006). Besides, as the cheese market faces intense international competition, the optimization of production processes becomes more critical for the economic success of dairy companies. For this reason, in some dairy productions, the cheese whey is frequently re-used to increase the yield or to improve the texture and nutritional value of the final cheese or produce high-value food products (Atamer et al., 2013).

Considering the two problematics exposed above about the lack of palative food options to supply daily ingest of protein in PKU patients and the environmental concern that implies the large quantities of cheese whey available as a byproduct, the object of this research was the production and expression of a recombinant Phe-free $\varkappa$-casein in K. lactis using cheese whey as carbon source, as well as its application as the protein source in a cream cheese formulation suitable for PKU patients consumption.

Theoretical Framework

Kluyveromyces Lactis

K. lactis is the most commonly used species belonging to the genus *Kluyveromyces* in the food industry. It can metabolize xylitol, cellobiose, and lactose in addition to pentose sugars containing xylose and arabinose (*Recombinant Protein Production in Yeasts*, 2012). *K. lactis* has been presented as one of the first efficient hosts constructed for homologous and heterologous protein production, in addition to its use as single-cell protein (SCP) (Spohner et al., 2016; *Recombinant Protein Production in Yeasts*, 2012). Approximately 100 heterologous proteins produced by *K. lactis* have been reported among which are the large-scale production of the native β-galactosidase and recombinant bovine chymosin (Spohner et al., 2016). *K. lactis* has advantages such as the feasibility of easy genetic manipulation, the existence of an easy-to-use reagent kit, the availability of a fully sequenced genome, a low amount of secreted endogenous proteins, and the presence of commercial enzymes from *K. lactis* having GRAS (generally regarded as safe) status by the FDA. Additionally, The prevalence of an oxidative metabolism provides this yeast with a particular advantage compared to other microorganisms since it allows the generation of high yields in biomass, which can be used as a source of nutrients to supplement animal and human diets (Rubio-Texeira, 2006). For such

reason, lactose metabolizing species of *Kluyveromyces* have been extensively used to convert cheese whey into valuable products (Spohner et al., 2016; Rubio-Texeira, 2006).

Use of *K. lactis* in cheese whey

The high-scale production of dairy products leads to the constant generation of vast amounts of a by-product known as cheese whey. Cheese whey has high biological and chemical oxygen demands, mainly owed to its 5% (w/w) contents in lactose. The lack of affordable methods for its elimination still forces industries to dump most of the residue into sewage and dried lands, a fact that poses a permanent hazard in terms of environmental pollution (Rubio-Texeira, 2006). The most economical way for industries to eliminate this residue is to convert it into commercially valuable products. High value-added products can be obtained from cheese whey through its fermentation, carried out by lactose-assimilating microorganisms. Despite the variety of bacteria, yeasts, and fungi available with such ability, only a few microorganisms such as K. Lactis has been approved as GRAS, and are therefore suitable for the conversion of this residue into products for pharmaceutical and food industries (Ryan & Walsh, 2016; Yadav et al., 2015; Rubio-Texeira, 2006).

Recombinant protein production by *K. lactis*

K. lactis can be used for either intracellular or secreted expression of proteins of interest. Secretion is the most common approach to protein expression in *K. lactis*, due to the product obtained is significantly pure and does not require severe lysis of yeast cells to isolate it. Figure 1 shows the secretion process of a recombinant protein through *K. Lactis.*

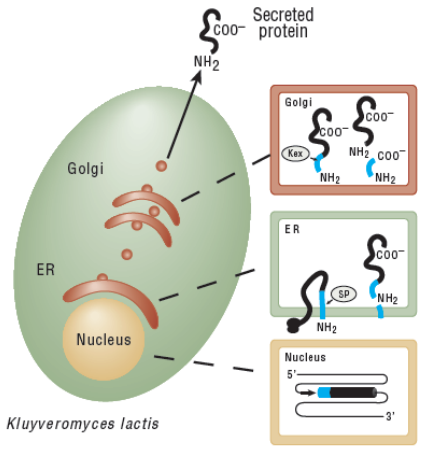

Figure 1: Protein secretion patway in *K. Lactis (NEB, 2007)*

Firstly, the DNA that encodes a fusion of proteins of the α-MF domain (blue) and a desired recombinant protein (black) is integrated into the nucleus. The expression is driven by the PLAC4-PBI promoter (identified as an arrow in Figure 1). Once the expression of the fusion protein has been performed, a signal peptide in the α-MF domain directs the translocation of the fusion protein into the endoplasmic reticulum (RE) and is removed by a peptidase (SP) signal. The secretory vesicles (orange circles in Figure 1) transport the fusion protein to the Golgi, where the Kex endoprotease cleaves the α-MF domain, releasing a mature form of the desired protein. Finally, the protein is transported through vesicles to the plasma membrane (PM), where it is secreted by the cell (NEB, 2007).

Currently there are commercial kits that provide methods and tools for cloning and expressing a gene of

interest in *K. lactis* by secretion using different strategies and vector pKLAC2 as it is described below (NEB, 2007; van Ooyen et al., 2006).

Cloning Strategy I, secretion of a protein with native N-terminus: Naturally secreted proteins contain native secretion leader sequences, in this case the forward primer must contain an XhoI site and a Kex protease cleavage site immediately followed by the first codon of the gene's open reading frame. The Reverse Primer must include a restriction site for cloning into any of the polylinker sites of pKLAC2. After Kex processing, the protein of interest containing a native N-terminus is secreted as shown in Figure 2 B(NEB, 2007).

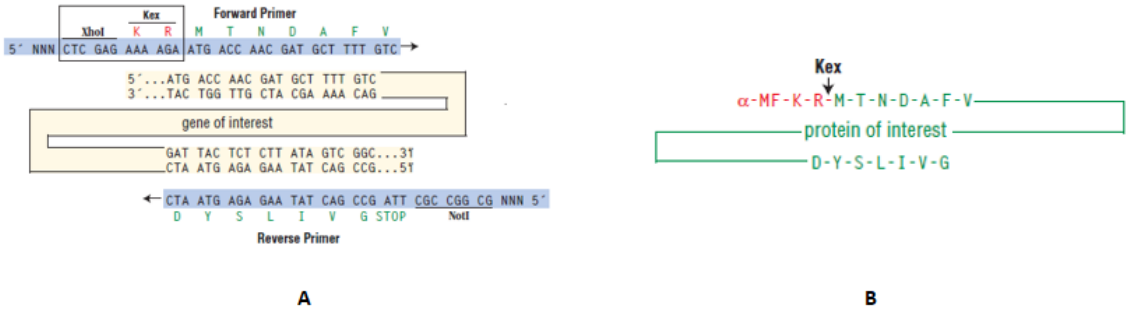

Figure 2: Cloning strategy I A) Gene of interest B)Protein of interest secreted (NEB, 2007)

Cloning Strategy II, secretion of a protein with non-native N-terminus: This case is when the gene of interest contains an XhoI site, the cloning of the gene may be in-frame with the α-MF domain using other sites in the pKLAC2 polylinker (NotI and BamHI). This strategy results in vector-encoded amino acids being added to the protein's N-terminus. After Kex processing, the secreted form of the protein of interest contains extra vector-encoded amino acids at its N-terminus (dashed underline in Figure 3 B)(NEB, 2007).

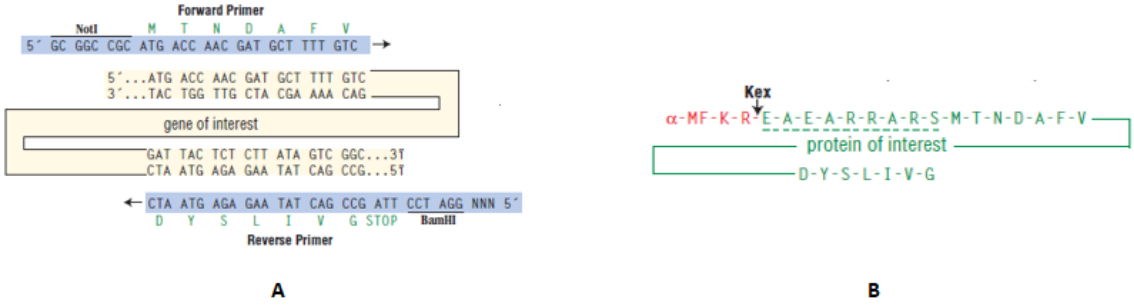

Figure 3: Cloning strategy II A) Gene of interest B)Protein of interest secreted (NEB, 2007)

Cloning Strategy III, secretion of a protein with a C-terminal hemagglutinin epitope tag: The Epitope tagging is used to allow for detection of secreted recombinant protein in the culture medium. It is recommendable to use PCR to incorporate DNA encoding a C-terminal hemagglutinin (HA) peptide epitope and a stop codon into the gene of interest(NEB, 2007).

Figure 4 A shows the gene of interest with a C-terminal antibody epitope tag (hemagglutinin epitope [HA] shown). The Forward Primer should be designed on the same way as Cloning Strategies I and the Reverse Primer should contain DNA encoding an HA epitope tag, a stop codon, and a restriction site for cloning into the polylinker (NotI). After Kex processing, the protein of interest containing a C-terminal HA-eptitope(underlined in Figure 3B) is secreted(NEB, 2007).

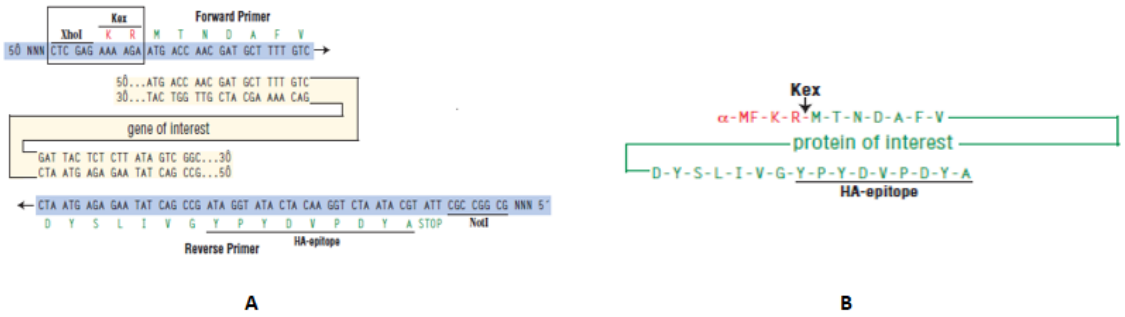

Figure 4: Cloning strategy III A) Gene of interest B)Protein of interest secreted (NEB, 2007)

For the production of Phe-free κ-casein, the objective of this research, it was decided to use the cloning strategy I described above since the secretion is preferable expression type due to its greater yields and reduction in purification costs (NEB, 2007; *Principles of Fermentation Technology (3rd Edition) - 12.2 Heterologous Protein Production by Bacteria*, 2017). Furthermore, this cloning strategy allows a native N-terminal protein similar to bovine k-casein.

Expression Vectors

The production of recombinant proteins in yeast commonly use "shuttle vectors" that can replicate in bacterial, typically *E. coli*, and also in yeast. This enables that the early stages of cloning are done in a bacterial system. Generally, these shuttle vectors are customized using as backbone a bacterial plasmid, then often include an E.coli origin of replication, and a bacterial selector marker (e.g., ampicillin resistance), a schematic representation of an expression vector is shown in figure 5 (*Principles of Fermentation Technology (3rd Edition) - 12.2 Heterologous Protein Production by Bacteria*, 2017; Greene, 2004).

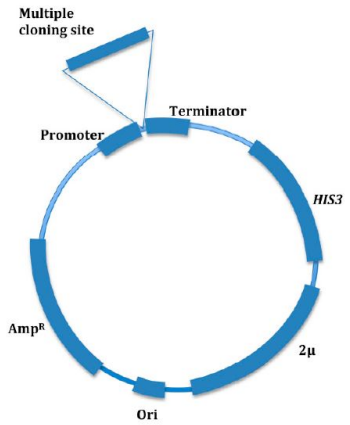

Figure 5: The schematic representation of a yeast expression vector. (*Principles of Fermentation Technology (3rd Edition) - 12.2 Heterologous Protein Production by Bacteria*, 2017)

There are two main types of vectors to produce heterologous genes in yeast for recombinant protein expression:

The first type is the episomal vector, which can auto-replicate, produce a high number of copies but needs to be made under the conditions that selection marker demand, otherwise they are lost (*Chapter 11: Heterologous protein expression in yeasts and filamentous fungi. In R. H. Baltz, A. L. Demain, & J. E. Davies (Eds.), Manual of industrial microbiology and biotechnology (3rd ed.)*, 2010). A common episomal vector for *K. lactis* is the plasmid pKD1, which is derivated from the natural S.cerevisiae 2 µm plasmid. This plasmid has been widely used for the production of several different recombinant proteins than include insulin, insulinase, Glyoxylate reductase, and Xylanase A latest with yields up to 100 g/L(Spohner et al., 2016).

The second one is the expression cassette, which is a linearized vector that contains a centromere sequence from the yeast genome, and the vector replication is linked to cell division. In comparison with episomal vectors, expression cassettes have a low number of copies, but higher stability, because there is no necessary the presence of selective environment to maintain the structure, which is an advantage in the large scale production process. It is essential to realize that the expression cassettes must be designed to be linearized before being incorporated into the yeast host (*Principles of Fermentation Technology (3rd Edition) - 12.2 Heterologous Protein Production by Bacteria*, 2017). In figure 6 is shown the schematic process of chromosomal integration of an expression cassette in *K. lactis*: here, the expression cassette, produced after an enzymatic linearization, is inserted in the LAC4 promoter region. Recombination with the K. lactis chromosomal DNA is possible due to the expression cassette is designed to comprise terminal parts which are homologous to sequences in the *K. lactis* genome (Spohner et al., 2016).

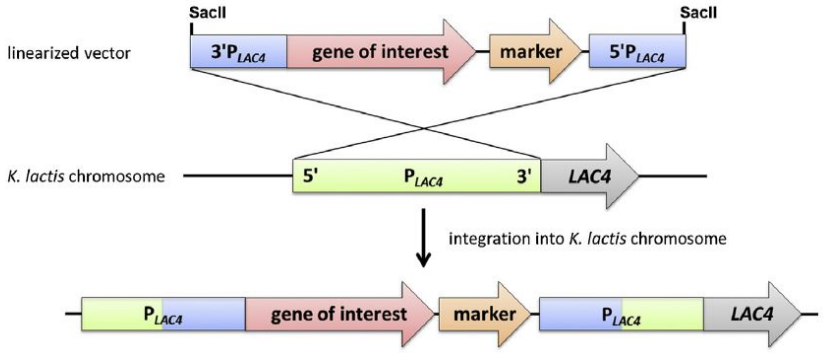

Figure 6: Schematic process of expression cassette chromosomal integration in *K.lactis (Spohner et al., 2016)*

The Integration of expression cassettes into the *K. lactis* genome is by insertion into the LAC4 promoter region. Here, an enzymatically linearized vector comprising terminal regions that are homologous to sequences in the *K. lactis* genome can be recombined with the chromosomal DNA. The most common expression cassette is the one obtained after the linearization of pKLAC2 plasmid provided by New England Biolabs (NEB) in the K. lactis Protein Expression Kit (NEB #E1000S) which offers a method for the secretion of recombinant proteins and is the typical approach to *K. lactis* expression (van Ooyen et al., 2006).

pKLAC2

In figure 7 is shown the map of pKLAC2 (9107 bp). The vector contains the native K. lactis mating factor leader sequence (noted as α-MF in figure 7); the role of the α-MF is to direct the protein of interest through the *K. lactis* secretory pathway.

The pKLAC2 vector is based on a LAC4 promoter variant that contains a mutated Pribnow box-like sequence (PLAC4-PBI) that has been engineered to lack background *E. coli* transcriptional activity. This let to clone

products in *E.coli*, even if they are toxic for unmodified *E. coli*, before their introduction in *K. lactis* cells. Once the plasmid is linearized by the enzymatic action of SacII or BstXI, the expression cassette resultant can integrate into the *K. lactis* genome at the LAC4 locus by homogeneous recombination (as shown in figure 6). The 5´ and 3´ ends of the LAC4 promoter (PLAC4-PBI) are separated by DNA encoding b-lactamase (ApR), which provide multi-resistance to β-lactam antibiotics, and the pMB1 origin (ori) to allow for its propagation in *E. coli*. Additionally, pKLAC2 contains a multiple cloning site (MCS) and the LAC4 transcription terminator (TT), which lies immediately downstream of 3´ PLAC4-PBI. Finally, the presence of the yeast ADH1 promoter (PADH1) drives the expression of an acetamidase selectable marker gene (amdS) (NEB, 2007; van Ooyen et al., 2006; Colussi & Taron, 2005).

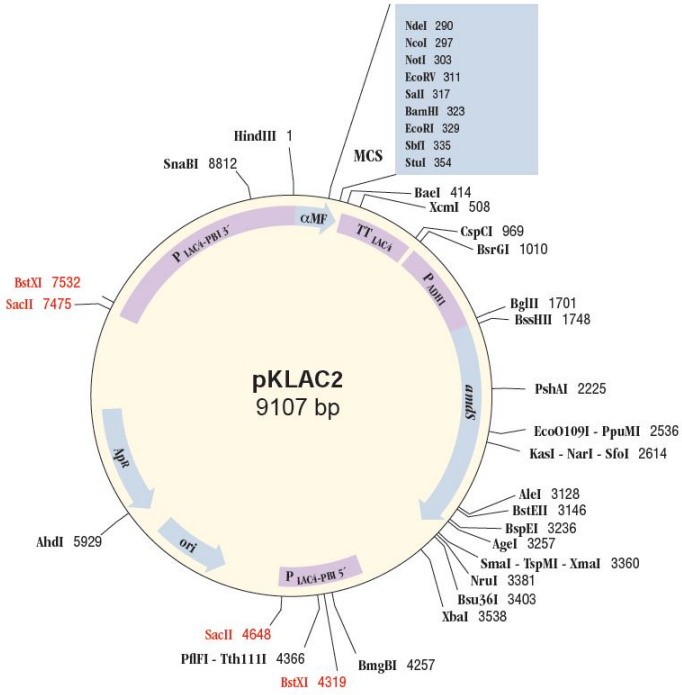

Figure 7: The pKLAC2 expression vector (NEB, 2007)

Plasmid production by *E. coli* and extraction with miniprep

The industrial enzymes or therapeutic recombinant proteins of commercial significance need to be produced in large quantities, and often issues of plasmid stability are critical for getting sustained production during the scale-up of cultures to industrial levels. Usually, microorganisms such as E. coli are routinely used for producing a recombinant gene of interest carried on a high copy number plasmid to boost gene dosage and, ultimately, boost process yield previously to its expression in the host microorganisms. (Walia et al., 2007; Cunningham et al., 2009).

On the other hand, in order to obtain a plasmid previously cloning with a gene of interest, miniprep, a molecular biology tool used by disrupting *E. coli* cell walls. For this, cells in the culture are disrupted with an alkaline shock, so that only their plasmid DNA is selectively infiltrated. A limitation is added with NaOH and SDS, which solubilizes the outer membrane of Gram-negative bacteria, releasing genomic DNA

and plasmid DNA, RNA, proteins, among others. Then, potassium acetate is added, which neutralizes all cellular waste and causes the plasmid DNA, which is in suspension, to renature rapidly, while the genomic DNA is precipitated by the high concentration of salts (especially SDS). By centrifugation, the supernatant containing the plasmid DNA is separated, and the RNA is removed with ribonucleases. Currently, there are column-based kits that facilitate miniprep in *E. coli* *(Qiagen, n.d.)*.

Proteins purification

Protein purification can be classified into two groups according to the expression form: purification of soluble and insoluble proteins. The classification is going to determine the strategy that can be carried out for its separation. Likewise, a microorganism can generate proteins that are mainly expressed in two ways: extracellular proteins, which means that they are secreted into the culture medium, or intracellular proteins, which are expressed inside the cell. In this last case, the protein could be located in the periplasm or in the cytoplasm("Estrategias de obtención de proteínas recombinantes en Escherichia coli", 2013).

Usually, extracellular proteins are more stable and smaller molecules compering with intracellular ones, and they can be purified using different physical separation methods. In contrast, the intracellular expressed proteins require the lysis of the cells that contain the protein; this can be achieved with different alternatives such as repeatedly freeze-thaw cycles, apply ultrasound waves, using high-pressure homogenization, or permeabilization with organic solvents, among others. The method is going to depend on the fragility of the protein and the resistance of the cell that would need to be broken (Lara, 2011).

In the case the protein was accumulated in the cytoplasm, the first step is cell disruption, which can be carried out using mechanical or chemical procedures. High-pressure homogenization is a scalable operation for this purpose. The accumulation of recombinant proteins in the cytoplasm frequently leads to the accumulation of inclusion bodies, which depends on factors such as the sequence of the protein, the strength of the promoter, temperature, growth rate, among others, but nowadays it is not easy to predict their formation. The formation of inclusion bodies is sometimes desirable (even fusion proteins are added to lead to their formation), since they protect recombinant proteins from proteolysis, and can be separated by continuous centrifugation (Lara, 2011).

Protein recovery can be carried out using filters of various sizes and with a charged surface. Filtration is also a scalable operation and is well established on the industrial scale. Nucleic acids can be removed by extractions or by adding polycationic agents such as polyethylamine. Protein can be selectively captured or precipitated using a variety of materials that have been developed for this purpose, such as some hydrophobic interacting resins (Kato et al., 2004). Final purification of the protein can be done by chromatographic methods such as anion and cation exchange, hydrophobic interaction and reverse phase chromatography (Lara, 2011).

Materials and Methods

Expression host strain and vector

The vector selected for recombinant Phe-free κ-casein Phe-free κ-casein production was the pKLAC2 due to is widely used for recombinant protein expression in *K. lactis*. The origin of replication of this vector (shown in figure 7) allow for its propagation in *E. coli* during cloning steps and the native *K. lactis* mating factor leader sequence (α-MF) lets the effective secretion of the Phe-free k-casein during expression in *K. lactis* as is described in . On the other hand, the organism used for Phe-free κ-casein Phe-free κ-casein expression is *K. Lactis* strain GG799 supplied together with pKLAC2 in the *K. lactis* Protein Expression Kit (NEB

#E1000S) from NEB. Additionally, it was used the plasmid pUC18 for Phe-free κ-casein gene synthesis. The plasmid maps were retrieved from the SnapGene tool.

The Phe-free K-casein gene (Phe-free CSN3) DNA sequence design

The design of Phe-free κ-casein gene (noted hereafter as Phe-free CSN3) were performed as follow:

Firstly, the Phe residues in the κ-*N Bos taurus* (bovine) amino acid sequence retrieved from Uniprot P02668 (Farrell et al., 2004; Holland & Boland, 2014) were replaced by Tyr resides, and the peptide signal (residues between Met_1 and Gln_{22}) removed to obtain a Phe-free κ-casein amino acid sequence.

Codon-optimized Phe-free CSN3

Secondly, the Phe-free κ-casein was back-translated using the Reverse Translate tool, which is based on the *E. coli* codon table from NCBI GenBank (Stothard, 2000; Benson et al., 2012). The resultant gene, Phe-free CSN3, was codon-optimized for expression in *K. lactis* using GenSmart Optimization tool. During the optimization process were eliminated the cleavage sites recognized by the restriction endonucleases XhoI, NotI, and SacII to avoid undesirable cleaves during the digestion and expression stages. After that, it was added the cut site bases recognized by Xhol and kex endonucleases at the beginning of Phe-free CSN3 sequence, and the NotI cut site at the end of the gene, following the cloning strategy for the secretion of a protein with native N-terminus recommended by the supplier (see section and figure 8).

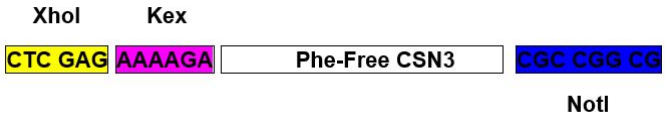

Figure 8: Phe-free CSN3 with XhoI, NotI and Kex cleavage sites.

Phe-free CSN3 virtual digestion

The DNA sequenced resultant was virtually synthesized in pKUC18 (Kang & Richardson, 1988) using Benchling. The pKUC18::Phe-free C3NS resultant was digested using Xhol and NotI at 37°C with the buffer solution 3.1 (supplied by NEB) that offers a 100% activity for both enzymes. The resultant fragments were analyzed through Sodium dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS–PAGE) also in Benchling. The SDS-PAGE results from Benchling are presented schematically in figure 9.

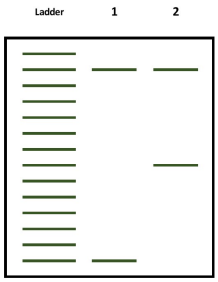

Figure 9: Schematic representation of SDS-PAGE results retrieved from Benching after virtual digestion.

Expression vector design and expression cassette production

The expression vector for Phe-free κ-casein production (noted hereafter as pKLAC2::Phe-free CSN3) was *in silico* designed through the digestion of Phe-free CSN3 and pKLAC2 with the single cutters restriction endonucleases XhoI and NotI provided also by NEB, and, which cut sites are in the MCS of pKLAC2. Xhol and NotI were selected following the cloning strategy for the secretion of a protein with native N-terminus as NEB recommends. After that, the pKLAC2::Phe-free CSN3 was enzymatic linearized with double cutter restriction enzyme SacII using buffer 2.1. at 37°C to produce the expression cassette.

Verification of Properly Integrated Cells with the expression cassette

It was used PCR in order to verify that the cells of *K. lactis* have correctly integrated the expression cassette for Phe-free κ-casein secretion. For this, it was designed two primers named integration primer 1 (forward) and integration premier 2 (reverse). The integration primer 1 was located in the LAC4 promoter region while integration primer 2 was located in the middle of Phe-free CSN3 (see figure 10). Both integration primers were designed procuring a length of 18 to 25 bp, a melting temperature about 50°C, and a GC content about 50% using Benchling wizard powered by Primer3 (Rozen & Skaletsky, 2000).

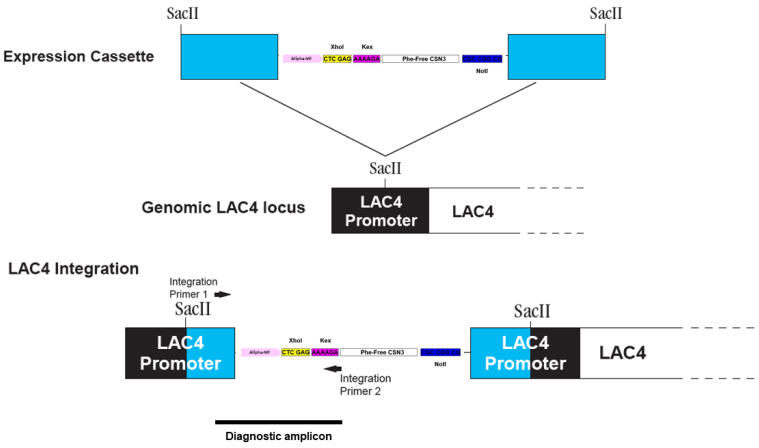

Figure 10: Representation of the strategy to verify proper integrated *K. lactis* cells using PCR based on supplier recommendations (NEB, 2007)

Results and Discussion

The Phe-free κ-casein gene (Phe-free CSN3) DNA sequence

The Phe-free κ-casein amino acid sequence used for reverse translation is shown in figure 11. The protein is 171 amino acid residues length after the removal of the peptide signal from the original κ-casein sequence.

```
        10         20         30         40         50
MQEQNQEQPI RCEKDERYYS DKIAKYIPIQ YVLSRYPSYG LNYYQQKPVA
        60         70         80         90        100
LINNQYLPYP YYAKPAAVRS PAQILQWQVL SNTVPAKSCQ AQPTTMARHP
       110        120        130        140        150
HPHLSYMAIP PKKNQDKTEI PTINTIASGE PTSTPTTEAV ESTVATLEDS
       160        170
PEVIESPPEI NTVQVTSTAV
```

Figure 11: The Phe-free κ-casein amino acid sequence without signal peptide

In figure 12 is presented the Phe-free CSN3 DNA sequence after reverse translation (513 bp). The analysis of cleavage sites in Phe-free CSN3 using Benchling shows that there is not cut sites recognized by XhoI nor NotI in the gene, but there is a cut site in the base 507 recognized by SacII that needs to be removed during the codon optimization to avoid problems during expression in *K. lactis* (see section).

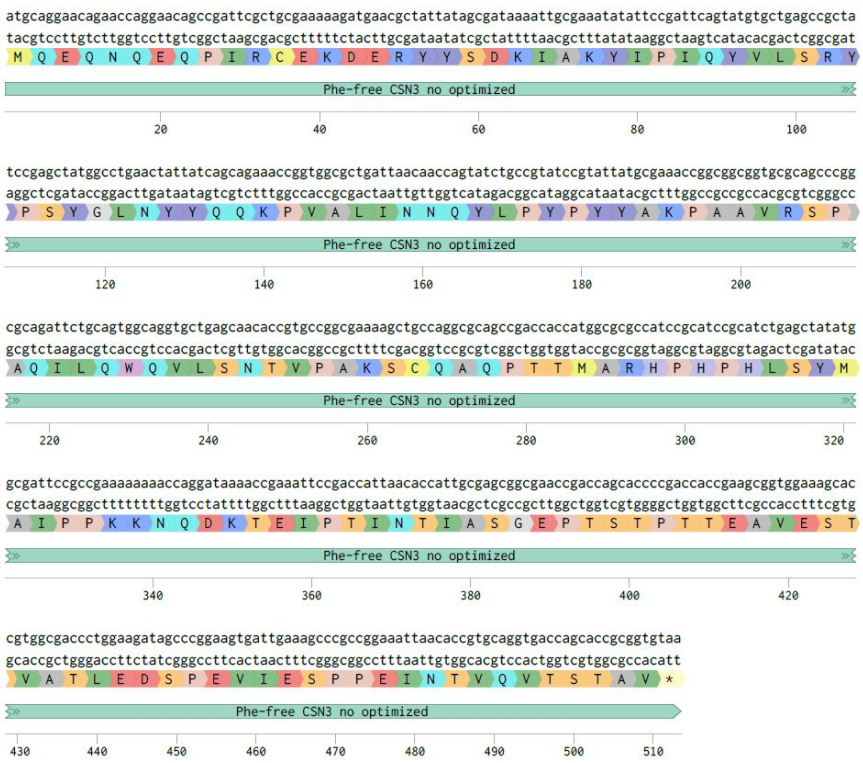

Figure 12: Phe-free CSN3 DNA sequence.

Codon-optimized Phe-free CSN3

The GC content comparison of original Phe-free CSN3 and the optimized version for expression in K. lactis is shown in table 1. It is observed that the codon-optimized Phe-free CSN3 has a lower GC content. According to (Sinclair & Choy, 2002), relatively low GC content is preferable to optimal expression of a mammalian gene in *K. lactis*. Futhermore, (Feng et al., 2009) aimed at a sevenfold increased level of chymosin production by the codon optimization for *K.lactis* expression of the encoding gen by decreasing the GC content from to 57.2%. to 32.9%. Under those circumstances, it is expected that codon-optimized Phe-free CSN3 allows a higher protein production level in comparison with original gene.

Gene	GC content	Codon bias organism
F-free CSN3 no optimized	55.17%	E. coli
Codon optimized F-free CSN3	42.69%	K. lactis

Table 1: Comparison of GC content in Phe-free $\varkappa$-CSN3 before and after optimization

On the other hand, codon optimization of Phe-free CSN3 effectively avoids the SacII cutting site initially presented in the gene DNA sequence, which is beneficial for the expression process as it was commented before. Figure 13 shows the final design of DNA sequence to be synthesized for Phe-free CSN3 production at which were added the XhoI and NotI cleavage sites. It is important to realize that between Xhol cleavage

site (noted in figure 13 in orange) and Phe-free CSN3 it was added a Kex endoprotease cut site (noted in yellow) that is necessary to produce a mature form of Phe-free κ-casein in the Golgi apparatus by cleaving the α-MF pro-domain as is explained in section (NEB, 2007; Curto et al., 2013). The expected length of Phe-free CSN3 is 526 bp.

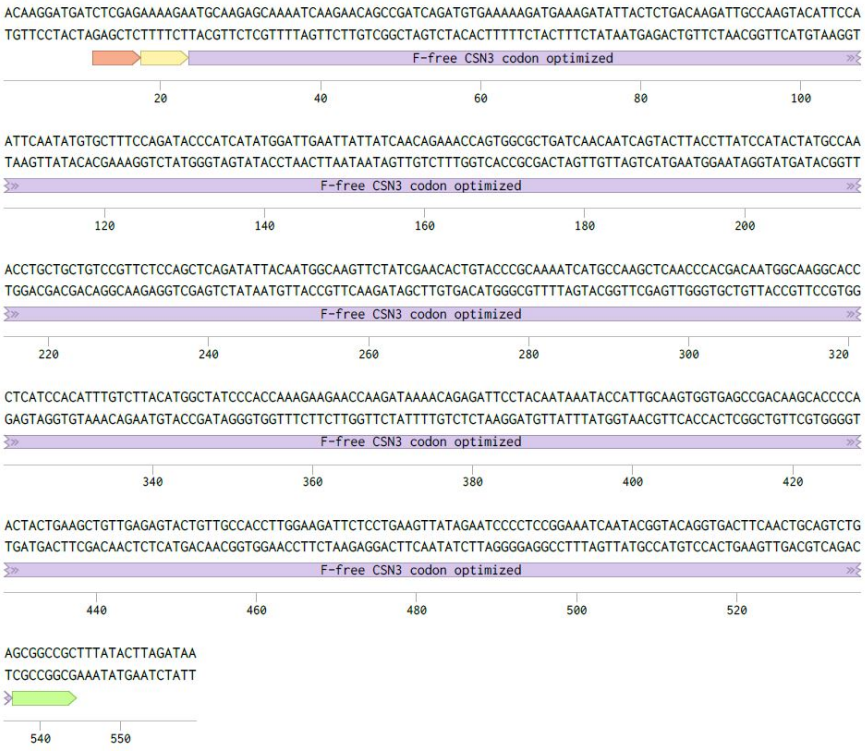

Figure 13: DNA sequence containing Phe-free CSN3 (purple), XhoI restriction site (orange), Kex (yelow), and NotI (green).

Phe-free CSN3 virtual digestion

The DNA sequence containing Phe-free CSN3 virtually synthesized into the XbaI-EcoRI cleavage sites in the MCS of pUC18 was analyzed through SDS-PAGE. The results are presented in figure 14. After virtual digestion with XhoI and NotI enzymes, it is obtained an SDS-PAGE with the ladder Life 1 kb Plus. The lane 1 presented two bands of 526 bp and 2.6 kbp that correspond to Phe-free CSN3 with Kex cut site and pUC18 residue respectively.

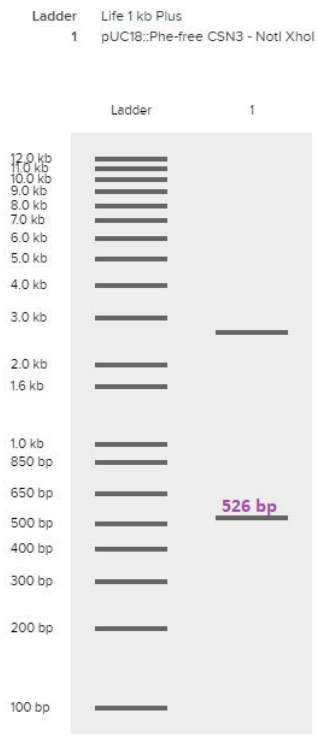

Figure 14: SDS-PAGE analysis for digested Phe-free CSN3. Band of 526 bp in lane 2 corresponds to Phe-free CSN3 digested with XhoI and NotI endonucleases.

Expression vector design and expression cassette production

The digested pKLAC2 with XhoI and NotI endonucleases is shown in figure 15, which is represented by the 9.1 kb band in the lane 2. The resulting vector (pKLAC2::Phe-free CSN3) after *in silico* enzymatic ligation is shown in figure 16. a SDS-PAGE analysis after digestion with the single cutter SacI at 37°C with buffer 1.1 results (lane 3) shows that the final length of pKLAC2::Phe-free CSN3 is 9.5 kb as is expected (see figure 15). Finally, the expression cassette obtained after enzymatic linearization of pKLAC2::Phe-free CSN3 with double cutter SacII is 6.7 kb size, as is shown in the land 4 of the SDS-PAGE analysis.

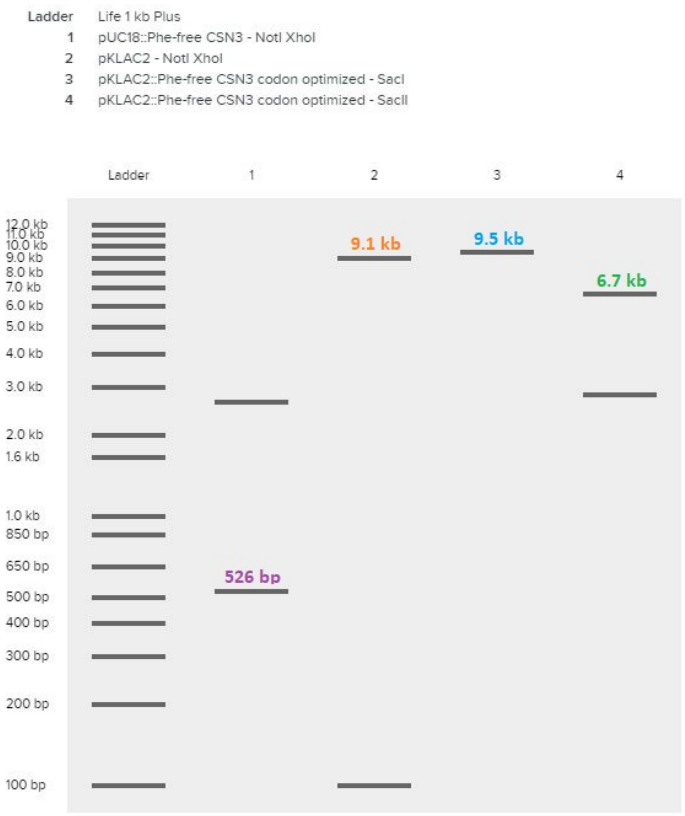

Figure 15: Cloning stages, SDS-PAGE analysis: Lane 1: digested Phe-free CSN3 ,with XhoI and NotI (526 bp). Lane 2: digested pKLAC2 with XhoI and NotI (9.1 kb). Lane 3: pKLAC2::Phe-free ϰ-CSN3 linearized with SacI (9.5 kb). Lane 4: expression cassette after enzymatic linearization with SacII.

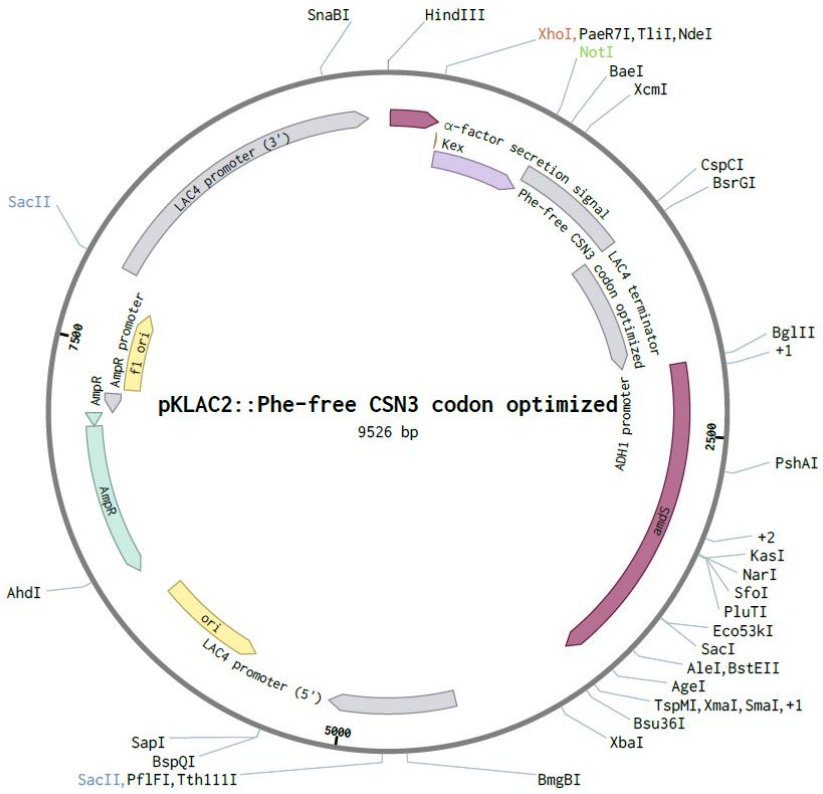

Figure 16: Codon-optimized pKLAC2::Phe-free κ-CSN3 . XhoI (orange), NotI (green) and SacII (blue) cutting sites are noted in the plasmid map.

Verification of Properly Integrated Cells with the expression cassette

The integration primers were designed using the Primer3, wizard function in Benchling procuring to avoid the region comprehend between 2418-6699 in the expression cassette (figure17) and a maximum GC content of 50%, a length between 18 to 25 bp and an optimal melting temperature of each primer of 58°C as is recommended by (Sambrook, 2001). The integration primers that best fit those criteria are shown in table 2.

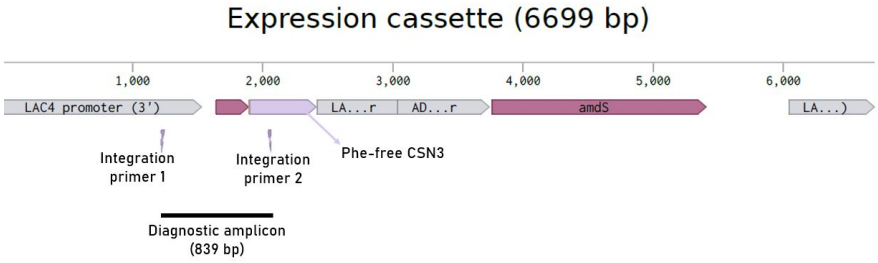

Figure 17: Expression cassette for Phe-free k-casein production in *K. lactis*

Name	Direction	% GC	T_m (°C)	Location	Length	Amplicon Product BP	Primer
Integration primer 1	Forward	0.5	58	1223-1244	22	839	5' TAGTGCGGTCGGTTACTTGGTT 3'
Integration primer 2	Reverse	0.5	58	2040-2061	22		5' TGATTGTTGATCAGCGCCACTG 3'

Table 2: Parameters of Integration primer 1 and 2 designed to verify chromosomal integration of Phe-free CSN3 in *K. lactis* cells

The designed primers allow a diagnosis amplicon of 839 bp. Secondary structure analysis of each integration primer at an estimated annealing temperature of 55°C shows that the following homodimer structure formed (figure 18).

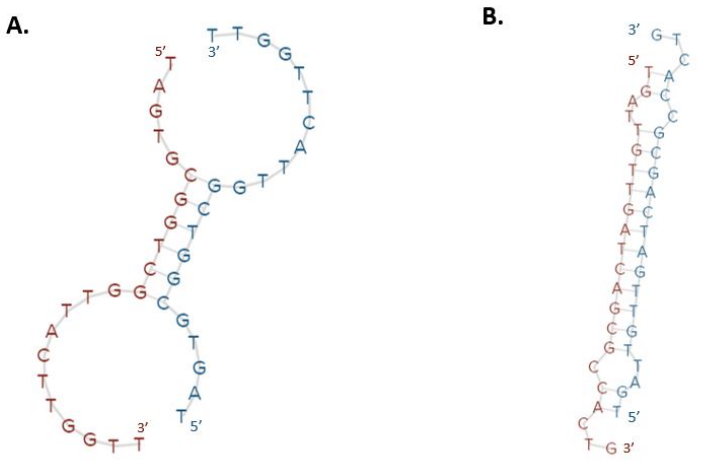

Figure 18: Primer secondary structure analysis. A. Homodimer formed in the Integration primer 1, Gibb's free energy: -2.86 kcal. B. Homodimer formed in the Integration primer 2, Gibb's free energy: -5-15 kcal.

As it is observed in the figure 18, there are in each homodimer structures some loops due to intermolecular interactions between base of the same primer. However, according to (Mukhopadhyay, 2018) a Gibb's free energy major to -6 kcal/mol is tolerable for the design of primers, since the intramolecular dimers are not much more readily than hybridizing to target DNA, then they do not reduce the product yield.

Recombinant Protein production and formulation of final product

Protein production

The first stage of the process is the production of Phe-Free κ-CSN3, for which the yeast *k. lactis* appropriately modified for this purpose is used, the growth medium will be mainly cheese whey. From nutritional view point, the average composition of whey is water (93%), lactose (4.5%–5.0% w/v), soluble proteins (0.6%–0.85% w/v), minerals (0.53%), and fat (0.36%) (Ryan & Walsh, 2016; Tesfaye et al., 2019), this way is possible to use as a carbon source for *k. lactis* fermentation the lactose in whey, as well as salts and small amounts of micronutrients that will allow the growth of *k . lactis* properly. Although this medium will provide the microorganism with an adequate environment for growth, the formulation of the complete broth must be optimized together with parameters as pH, temperature and dissolved oxygen for fermentation, the latter being fundamental in the growth of yeasts. According to (Merico et al., 2009) the fermentation can be carried out at 30 ° C, pH 5.0 controlled with

The cultivation of this microorganism will end up with the production of the extracellular Phe-Free κ-CSN3, then protein secreted into the culture medium should be separated or purified in order to be used.

Protein Purification

Proteins expressed in the extracellular space from the culture media of bacteria, fungi, cells of animals or plants, generally do not contain a large number of contaminating proteins, the protein of interest is the going to be the most abundant, especially if it is produced as a recombinant protein. However, the protein can be very diluted in the initial sample and large volumes are required to be processed and obtain a higher concentration of the protein (Rueda, 2010).

Having the *k.lactis* culture with the Phe-Free K-casein excreted in the broth at its best point of the growing curve, the next step is to separate the microorganism and other impurities that are going to be present at the end of fermentation. For this the best option is used the ultrafiltration process with adequate Dalton membrane size to achieve separate the protein of interest (C.A. Solís, 2017). However, is necessary to make some previous process before to eliminate other components.

As the protein is not inside the microorganism, the first step is to perform a centrifugation, this allows to separate the biomass present in the broth, from the supernatant where the proteins are going to be located. After this, a filtration with large pore size is recommended (around 2 um), to eliminate the solid particles of great size, and sever other impurities of larger sizes that could affect the ultrafiltration process, that could be followed but other filtration process with smaller membrane (around 0.2 um). This size should be define with specificity having the optimize broth formulation.

Finally, the ultrafiltration process is going to carry out, for this is necessary to use the adequate membrane. Phe-Free κ-CSN3 was modified replacing a type of nucleic acids of bovine k-casein; thus we can assume that the size of regular kappa casein is the same for this recombinant protein. Having that, the theoretical molecular weight of k-casein is reported as 18.974 Da (Holland, 2008). This process could be perform using continuous or multi-stage design to optimize the use of the membrane and guarantee the best recovery of the protein. (C.A. Solís, 2017)

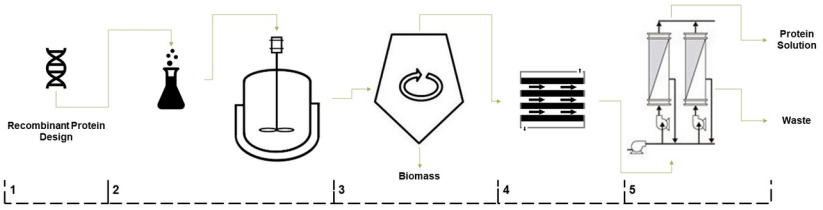

Figure 19: Production and purification process. 1. Recombinant protein design/cloning. 2. Fermentation from inoculum to bioreactor of the modified k.lactis to produce the recombinant protein. 3. Continuous centrifugation process, separates the biomass from growth medium with protein content. 4. Tangential filtration to separate the bigger broth impurities. 5. ultrafiltration for separate only the protein of interest. .

Product formulation: a cream cheese

Once we have purified protein (Phe-Free K-CN) is possible to purpose a manufacture scale operation where the protein is going to be used in the formulation of cream cheese. The K. lactis, yeast has a reported performance between 10 to 50 mg/L of produced protein (NEB, 2007), although this could be optimize with a good formulation and process conditions to obtain greater results, the average performance will be around the tens milligrams per liter of culture to take into account in the scale up process.

On the other hand, production data from milk manufacture reports that the whey production represent 211500 kT/for year 2020, therewith we will be able to build to simulation of production of cream cheese(Fischer & Kleinschmidt, 2015).

Cream cheese manufactory

Cream cheese products are often categorized into two main types based on the different fat content in the initial mix and the final composition. These are double-cream cheese with at least 9-11% fat content in the initial mix, and single-cream cheese with 4.5-5% fat content in the initial mix, however there are also other similar kinds of cream cheeses based on different fat and dry matter contents (FDA regulations state that cream cheese has to have at least 33% fat and not more than 55% moisture content) (Chandan, 2003; Phadungath, 2005).

A common chemical composition of cream cheese is presented in figure 20, as the regular process of fermentation from milk is not possible to carry out in this case because of the different protein content it has and their effect on Phenylketonuria patients, the fermentation should perform using directly a broth with lactose, fat from a different source, salts and micronutrients, and all the protein content will be the purified Phe-Free $\varkappa$-CSN3 .

	Moisture	Fat	Protein	Lactose	Salt
Double Cream Cheese	60	30	8-10	2-3	0.75
Cream Cheese	70	14	12	3.5	0.75

Figure 20: Chemical composition (%w/w) of cream cheese (Phadungath, 2005)

Simulation of production cream cheese

Manufacturing Procedure of cream cheese involves mainly the fermentation process, normally can be carry out using microorganisms as *Lactococcus lactis ssp.*, *Leuconostoc cremoris* or *Lactobasillus casei* among others (Chandan, 2003). In this case the purpose is to use the *K. lactis,* a yeast that naturally metabolize the lactose in an acid-lactis fermentation. The process is normally carry out for several hours until the moment the cheese is at the isoelectric point, usually a pH of 4.47 - 5.81 is attained, for this specific case, the modified proteins is going to be the one provides all the protein content to the broth, and as the Phe-Free k-casein have a greater isoelectric point (5.703), a higher pH is enough, therefore the fermentation process is supposed to take less time as the regular one.

For the fermentation process, as it is not possible to use regular milk, it will be necessary use as a growing broth for the yeast, a mixture adding water, fat, protein (only the Phe-free K-Casein produced), lactose and salt in accordance to the reported portions usually used in this type of products (figure 20). First able, the water should add for intervals of time and temperature (130 C for 5 min, after, 105 for 1 min and 105 C). Because, these methodology help to guarantee humidity for cream cheese, after add fat (canola oil) and protein, therefore stirring start slowly and gradually accelerate until you have reached 2,000 rpm. Finally, these should mix lactase and salt for ensure quality as sensory evaluation of cream cheese.(Ningtyas et al., 2017)

Knowing that the sales of cream cheese in Colombia represent 5% of 54 million kilograms cheese consumption during the 2018-2019 (Portafolio, 2019), and its important have to consideration of yeast performance. In this vein, we will project to produce 550 Kg per turn and represent as:

Kind of cheese	Colombia (mount)	Unit	Porcent
Cheese	54000000	Kg/year	100%
Cream Cheese	2700000	Kg/year	5%
Mount of production	27000	Kg/year	1%
Mount per mount	2250	Kg/year	-

Proceso	Mount of whey	Unit
Whey (2020)	211500	Kton/year
Mount In	30599	KG
Restriccion	2	Horas de recidencia
Production de K-casein	0,00006	Kg/L
Production of K-casein per turn	7	Kg/ per turn

Manufacturing of cream cheese		Mount of production (KG)
Ingredients	%	25
Fat (Canola oil)	14%	3,5
Water (muisture)	64%	16
Protein (K-CN)	12%	3
Lactase	3%	0,75
Salt	7%	1,75

Producton per monuth	1800	Kg/mounth

Figure 21: Scale of manufacturing plan for cream cheese

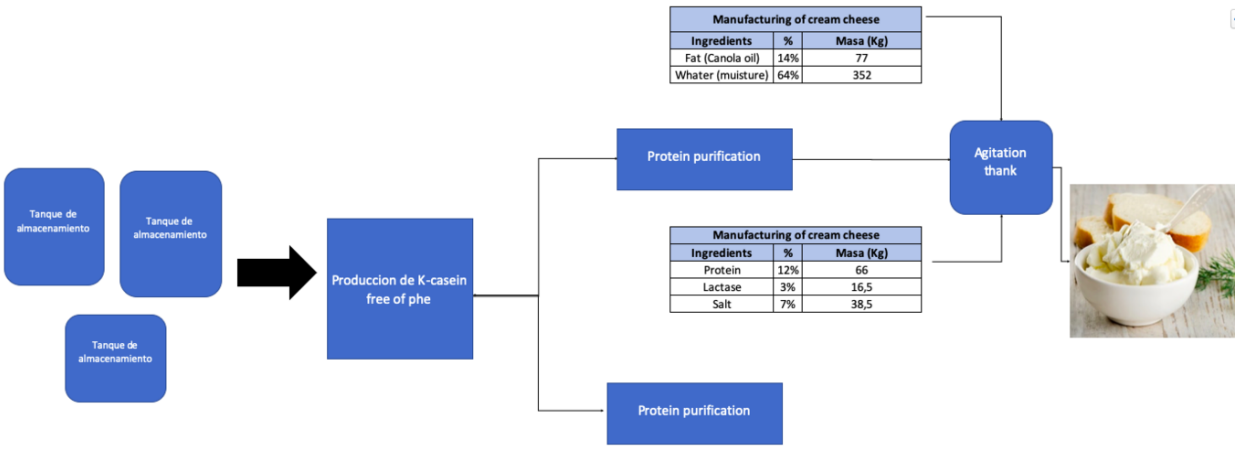

Figure 22: Diagram of process for cream cheese per production volume

Conclusions

- The *in silico* design of a suitable gene for the production of Phe-free κ-casein using and expression cassette built from pKLAC2 for its optimized secretion in *K. lactis* was possible due to the use of powerful tools as Benchling, Snap gene, Reverse Translate, and GenSmart Optimization. Additionally, a pair of suitable Integration primers for verification of properly integrated *K. lactis* cells with the expression cassette containing the Phe-free κ-casein gene was possible using Primer3 powered by Benchling.
- It was proposed a feasible scale-up procedure for the purification of Phe-free κ-casein and its use in a cream cheese formulation suitable for PKU patients.
- Finally, it was proposed a formulation for Phe-free cream cheese, including 12%wt. of Phe-free k-casein as the protein source; 14% at. of canola oil, as grease source; 3% wt. of lactose for lactic acid production during acid fermentation and, and as a sugar source; and 7% at and water. of salt. It is expected to produce 27,000 kg of cream cheese per year to address 1% of the current Colombian market for cream cheese.

References

Molecular Cloning and Expression of Bovine -Casein in Escherichia coli. (1988). *Journal of Dairy Science*, *71*(1), 29–40. https://doi.org/10.3168/jds.s0022-0302(88)79521-1

Genetic engineering of bovine .kappa.-casein to improve its nutritional quality. (1991). *Journal of Agricultural and Food Chemistry*, *39*(2), 422–427. https://doi.org/10.1021/jf00002a040

Cloned transgenic cattle produce milk with higher levels of -casein and -casein. (2003). *Nature Biotechnology*, *21*(2), 157–162. https://doi.org/10.1038/nbt783

Alterations of the Physical Characteristics of Milk from Transgenic Mice Producing Bovine -Casein. (1996). *Journal of Dairy Science*, *79*(5), 791–799. https://doi.org/10.3168/jds.s0022-0302(96)76427-5

De cada 100 mil nacimientos entre dos y seis bebés pueden desarrollar la fenilcetonuria. (2010). https://www.cronicadelquindio.com/noticia-completa-titulo-de_cada_100_mil_nacimientos_

entre_dos_y_seis_bebes_pueden_desarrollar_la_fenilcetonuria-seccion--nota-23103.htm. https://www.cronicadelquindio.com/noticia-completa-nota-23103

(2007). *Journal of Biotechnology*, *128*(2), 383–392. https://doi.org/10.1016/j.jbiotec.2006.10.016

Recombinant expression analysis of natural and synthetic bovine alpha-casein in Escherichia coli. (2000). *Applied Microbiology and Biotechnology*, *54*(5), 671–676. https://doi.org/10.1007/s002530000435

Improvement cloning, and expression of an in silico designed protein enriched with large neutral amino acids in Pichia pastoris for possible application in phenylketonuria. (2020). *Journal of Food Biochemistry*, *44*(3). https://doi.org/10.1111/jfbc.13151

Phenylketonuria. (2019). *Paediatrics and Child Health*, *29*(3), 111–115. https://doi.org/10.1016/j.paed.2019.01.001

Food products made with glycomacropeptide, a low-phenylalanine whey protein, provide a new alternative to amino Acid-based medical foods for nutrition management of phenylketonuria.. (2012). *J Acad Nutr Diet*, *112*, 1201–1210.

Kluyveromyces lactis: An emerging tool in biotechnology. (2016). *Journal of Biotechnology*, *222*, 104–116. https://doi.org/10.1016/j.jbiotec.2016.02.023

(2012). Humana Press, Totowa, NJ.

The biotechnological potential of whey. (2016). *Reviews in Environmental Science and Bio/Technology*, *15*(3), 479–498. https://doi.org/10.1007/s11157-016-9402-1

Cheese whey: A potential resource to transform into bioprotein, functional/nutritional proteins and bioactive peptides.. (2015). *Biotechnol Adv*, *33*, 756–774.

Endless versatility in the biotechnological applications of Kluyveromyces LAC genes.. (2006). *Biotechnol Adv*, *24*, 212–225.

Dairy and Nondairy-Based Beverages as a Vehicle for Probiotics Prebiotics, and Symbiotics: Alternatives to Health Versus Disease Binomial Approach Through Food. (2019). In *Milk-Based Beverages* (pp. 473–520). Elsevier. https://doi.org/10.1016/b978-0-12-815504-2.00014-1

Review: elimination of bacteriophages in whey and whey products. (2013). *Frontiers in Microbiology*, *4*. https://doi.org/10.3389/fmicb.2013.00191

K. lactis Protein Expression Kit: Instruction manual NEB E1000S. (2007).

Heterologous protein production in the yeastKluyveromyces lactis. (2006). *FEMS Yeast Research*, *6*(3), 381–392. https://doi.org/10.1111/j.1567-1364.2006.00049.x

(2017). Elsevier.

Host Cell Compatibility in Protein Expression. (2004). In *Recombinant Gene Expression* (pp. 003–014). Humana Press. https://doi.org/10.1385/1-59259-774-2:003

(2010). ASM Press.

Kluyveromyces lactis LAC4 Promoter Variants That Lack Function in Bacteria but Retain Full Function in K. lactis. (2005). *Applied and Environmental Microbiology*, *71*(11), 7092–7098. https://doi.org/10.1128/aem.71.11.7092-7098.2005

Development of expression vectors for Escherichia coli based on the pCR2 replicon. (2007). *Microbial Cell Factories*, *6*(1), 14. https://doi.org/10.1186/1475-2859-6-14

Factors affecting plasmid production in Escherichia coli from a resource allocation standpoint. (2009). *Microbial Cell Factories*, *8*(1), 27. https://doi.org/10.1186/1475-2859-8-27

https://www.qiagen.com/us/. http://www.qiagen.com/us/

(2013). *VacciMonitor*. http://scielo.sld.cu/pdf/vac/v22n2/vac06213.pdf

Producción de proteínas recombinantes en Escherichia coli. (2011). *Revista Mexicana De Ingeniería Química*, *10*. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-27382011000200006

Hydrophobic interaction chromatography at low salt concentration for the capture of monoclonal antibodies. (2004). *Journal of Chromatography A*, *1036*(1), 45–50. https://doi.org/10.1016/j.chroma.2004.02.009

Nomenclature of the Proteins of Cows' Milk—Sixth Revision. (2004). *Journal of Dairy Science*, *87*(6), 1641–1674. https://doi.org/10.3168/jds.s0022-0302(04)73319-6

Post-translational Modifications of Caseins. (2014). In *Milk Proteins* (pp. 141–168). Elsevier. https://doi.org/10.1016/b978-0-12-405171-3.00005-2

The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences. (2000). *BioTechniques*, *28*(6), 1102–1104. https://doi.org/10.2144/00286ir01

GenBank. (2012). *Nucleic Acids Research*, *41*(D1), D36–D42. https://doi.org/10.1093/nar/gks1195

Primer3 on the WWW for General Users and for Biologist Programmers. (2000). In *Bioinformatics Methods and Protocols* (pp. 365–386). Humana Press. https://doi.org/10.1385/1-59259-192-2:365

Synonymous codon usage bias and the expression of human glucocerebrosidase in the methylotrophic yeast Pichia pastoris. (2002). *Protein Expression and Purification*, *26*(1), 96–105. https://doi.org/10.1016/s1046-5928(02)00526-0

Codon optimization of the calf prochymosin gene and its expression in Kluyveromyces lactis. (2009). *World Journal of Microbiology and Biotechnology*, *26*(5), 895–901. https://doi.org/10.1007/s11274-009-0249-2

Establishing the Yeast Kluyveromyces lactis as an Expression Host for Production of the Saposin-Like Domain of the Aspartic Protease Cirsin. (2013). *Applied and Environmental Microbiology*, *80*(1), 86–96. https://doi.org/10.1128/aem.03151-13

Molecular clonign: a laboratory manual. Chapter 8. (2001). Cold Spring Harbor Laboratory Press.

Chapter 17. Primer design basics from Basic applied bioinformatics. (2018). John Wiley & Sons.

The oxygen level determines the fermentation pattern inKluyveromyces lactis. (2009). *FEMS Yeast Research*, *9*(5), 749–756. https://doi.org/10.1111/j.1567-1364.2009.00528.x

Expresión y purificación de las proteínas estructurales del rotavirus VP5 y VP8* en bacterias E. coli BL21(DE3) — Moreno — Revista Colombiana de Biotecnología*. (2010). [Master's thesis, https://revistas.unal.edu.co/index.php/biotecnologia/article/view/39841/41966]. https://revistas.unal.edu.co/index.php/biotecnologia/article/view/39841/41966

Membrane technology: Ultrafi ltration. (2017). *Entre Ciencia e Ingeniería*, *22*. http://www.scielo.org.co/pdf/ecei/v11n22/1909-8367-ecei-11-22-00026.pdf

Post-translational modifications of caseins. (2008). In *Milk Proteins* (pp. 107–132). Elsevier. https://doi.org/10.1016/b978-0-12-374039-7.00004-0

Synthesis of galactooligosaccharides using sweet and acid whey as a substrate. (2015). *International Dairy Journal*, *48*, 15–22. https://doi.org/10.1016/j.idairyj.2015.01.003

CHEESES | Soft and Special Varieties. (2003). In *Encyclopedia of Food Sciences and Nutrition* (pp. 1093–1098). Elsevier. https://doi.org/10.1016/b0-12-227055-x/00201-7

Cream cheese products: A review. (2005). *Songklanakarin Journal of Science and Technology*, *27*. https://rdo.psu.ac.th/sjstweb/journal/27-1/18cream-cheese.pdf

Texture and lubrication properties of functional cream cheese: Effect of -glucan and phytosterol. (2017). *Journal of Texture Studies*, *49*(1), 11–22. https://doi.org/10.1111/jtxs.12282

Consumo de queso sigue creciendo en Colombia. (2019). *Business, Portafolio*, *https://www.portafolio.co/negocios/consumo-de-queso-sigue-creciendo-en-colombia-530645*.

Reverse Translate. https://www.bioinformatics.org/sms2/rev_trans.html. https://www.bioinformatics.org/sms2/rev_trans.html

Home. https://www.qiagen.com/us/. http://www.qiagen.com/us/

QIAGEN. https://www.qiagen.com/us/. http://www.qiagen.com/us/

Chandan, J. S., Okoth, K., Gokhale, K. M., Bandyopadhyay, S., Taylor, J., & Nirantharakumar, K. (2020). Increased Cardiometabolic and Mortality Risk Following Childhood Maltreatment in the United Kingdom.. *J Am Heart Assoc*, e015855.

Chandan, J. S., Taylor, J., Bradbury-Jones, C., Nirantharakumar, K., Kane, E., & Bandyopadhyay, S. (2020). COVID-19: a public health approach to manage domestic violence is needed.. *Lancet Public Health*.

Lučan, M., Čačić, J. L., Jukić, M., Obrdalj, H. T., Gajari, D., Primorac, L., Cvetković, T., Slačanac, V., & Ranilović, J. (2020). Physico-chemical properties spreadability and consumer acceptance of low-sodium cream cheese. *Mljekarstvo*, *70*(1), 13–27. https://doi.org/10.15567/mljekarstvo.2020.0101

Stressler, T., Leisibach, D., Lutz-Wahl, S., Kuhn, A., & Fischer, L. (2016). Homologous expression and biochemical characterization of the arylsulfatase from Kluyveromyces lactis and its relevance in milk processing. *Applied Microbiology and Biotechnology*, *100*(12), 5401–5414. https://doi.org/10.1007/s00253-016-7366-2