Allergen modification of the ovalbumin as possible method to minimize the egg allergenic reaction into the human organism and its inclusion in a food product

Yiovann Alirio Arce Portilla[1], Felipe Rojas[1], Natalia Erazo Clavijo[1], Jessica Agresott[1], and Daniela Guáqueta[1]

[1]University of Los Andes (Uniandes)

May 22, 2020

Abstract

The production strategy for enhanced ovalbumin protein was realized based on the allergenic properties described in the previous delivery. As a production agency *Pichia Pastoris* was established which is a highly used eukaryotic organism for the mass production of proteins and *pPIC9K* was chosen as the cloning vector of the sequence to be synthesized. For the protein production strategy, the computer tools available on the web; *sequence Manipulation Suite* (www.bioinformatics.org) were used for reverse translation of the protein; *Codon Usage Database* (www.kazusa.or.jp) for the most likely codon usage table in *Pichia Pastoris*; *EMBL* (www.embl.de) for vector selection of interest; *SnapGene* (www.snapgene.com) as a search repository for plasmid mapping and *Benchling* (www.benchling.com) for the gene cloning process to be synthesized with the optimized codon sequence for our ovalbumin protein.

1. Introduction

Nowadays into a balanced daily diet two needs arise and consist in include functional and nourishing foods (Bhat & Bhat, 2011). However, currently there is a significant population that suffers certain limitations when consuming some foods due to they generate an allergic reaction (Sánchez & Sánchez, 2015). Within the market are products where it is excluded those allergic ingredients and they are substituted by other ones that have some similarities such as vegan food which no contains milk either egg. Also, these two are mostly affect to children with 5 -8 years old, and they tend to be outgrown (EUROPE, 2013). This trend was even coupled in other industries such as cosmetics and nutraceuticals which are named as hypoallergenic (M. Balsam, 2009). There are no clear proposals that aim to promote the consumption of these food into a final product, even without any improvement that has been achieved with biomolecular science modifications to reduce their allergic effect.

One of those daily foods that most of the people consume is egg. This is constituted by 55% of egg white (Rc et al., 2018), and contain certain proteins that cause allergy. The ovalbumin composes around of 65% of white, and is characterized by a sequence of amino acids that generate this allergic reaction (de Estudios del Huevo, 2009). The database created by National Institute of Allergy and Infectious Diseases reports almost the total ovalbumin chain is allergic, defining 298 epitopes as possible source of allergy. In 2003, (Mine & Rupa, 2003), determined along the chain there are five epitopes as the several sites that promotes the allergy. When the body tries to digest the entire protein, it recognizes like a threat, thus the body activates the antibody agents as a red flag. This causes allergic symptoms such as a rash, redness on the skin, and even closure of the trachea, reducing the ability to breathe (EUROPE, 2013).

Ovalbumin is a phosphoglycoprotein constituted by 368 amino acids and is generated in the oviduct of avian animals due to the located glandules in "magnum", which secretes the constituents of the egg white. It is part of the serine proteins, but its main characteristic is not to inhibit proteases, as this type of protein generally does (Huntington & Stein, 2001). Its real function in avian is not clear till this moment. However, it is consumed due to its high content part into the egg white and it was demonstrated that it can support some functions the serum albumin accomplishes in the human body. For instance, the transport of physiological and non-physiological substances that are not soluble in water, the maintenance of oncotic pressure, and the contribution of buffering capacity of the blood and even maintain albumin levels in people who have chronic kidney disease.

Based on the first report of the current project, which was determined the opportunity of the design of a product based on inclusion of a modified ovalbumin into a food daily product. This modification was developed with the exchange of some similar amino acids according with their hydrophobicity, and that are located on those five epitopes that were reported by (Mine & Rupa, 2003). The corresponding modifications showed that the tertiary conformation of the protein was maintained with respect to the original.

Also, the prediction of modified and original protein location showed a high probability of being extracellular as well as transmembrane location in 30-50 amino acid positions. The first indicates the molecule have hydrophilic radicals (Karp, 2011), and the second involves fulfilling functions such as being receptors that recognize and fix some molecules called ligands, channels or transporters of ions and solutes through the membrane and identifying nutrients, hormones or neurotransmitter (Karp, 2011). Finally, with this proposal the proteins structure was preserved which indicates the active sites are expected the same and it could be accomplished as the enhanced protein that will be added in a food product with a decrease in its allergic effects.

In this part of the project is proposed the microorganism and its production pathway to guarantee the fulfillment of having a modified ovalbumin produced and included into a daily product. Therefore, there was considered as *Pichia Pastoris* as the adequate microorganism to produce it based on recombinant technology. This consisted in determining the coding DNA sequence, the sequence expressed in codons, the RBS suitable for expressing GALNS, and making the cloning Fdesign.

2. Theoretical framework

The heterologous or recombinant protein production depends on the host to choose to produce the protein interest. This process consists in the isolation of the gene and subsequently cloned in an expression vector (Clark & Pazdernik, 2015). There are different organisms which act as the host such as animal cells, yeast, bacteria or fungus. However, their performance will depend on the preference in which codons are formed for each amino acid (Clark & Pazdernik, 2015). The protein expression system could be prokaryotic or eukaryotic. Prokaryotic is the renamed *Escherichia coli*, however this bacterium as prokaryotic cell microorganism has some limitations when a heterologous protein is going to be produced (García-Fruitós et al., 2014). Eukaryotic cells facilitate the postranslational modification process, like the glycosylation, for the proteins that need this. Therefore, the cultures of yeast, insect cells, microalgae, filamentous fungus, and mammalian cells can handle it(Clark & Pazdernik, 2015; García-Fruitós et al., 2014).

Human proteins are characterized for their postranslational modifications such as glycosylation, phosphorylation, acylation, carboximetilation, and others (Galindo & Ramirez, 1998). Most of the therapeutic proteins that are in the market were obtained through a eukaryotic system that permit those protein modifications(García-Fruitós et al., 2014). Mammalian cells is the first choice due to its ability to achieve adequate needed modifications, however, its maintenance is difficult due to it is very susceptible with environment changes and the treatment that they are exposed. In the last decade, yeast is the next option that which has been implemented to the recombinant human proteins, and the first selected fungus strain was the *Saccharomyces cerevisiae* in 1980, but some works affirm this has a unstable plasmid, low yield recorded, and hyperglycosilation phenomena (Rivero et al., 2016; Allende, 2007). In contrast, *Pichia*

pastoris a methylotrophic yeast that has demonstrated its high yield production in the proper form for those therapeutic molecules. Due to its feeding source (methanol) most of the expression system are based on inducible promotes for methanol. Some of those recombinant proteins that have been cloned and expressed in this microorganism are: Kunitz inhibitor proteases, C fragment of tetanic toxin, human interleukin II, human lysozyme, human serum albumin, and among others (Rivero et al., 2016).Its high quality of performance is supported by the respiratory metabolism allows to be achieved enormous cellular densities in the bioreactors (Chen et al., 2012).

2.1. Pichia Pastoris

 Pichia Pastoris is a single-celled eukaryotic organism called methylotophan yeast, which has a particular taxonomy, such that the genus *Pichia* is part of the family *Saccharomycetaceae*, the order *Saccharomycetales*, the Class *Saccharomycetes*, the *Ascomycota* phyleus and the *Fungis* kingdom, which is characterized by forming creamy, white and well-defined colonies. It is a highly used yeast as an expression system for the production of recombinant proteins, both for basic research and industrial purposes, this due to its easy genetic manipulation, its high levels of intra and extracellular production of the protein of interest and its ability to make post-translational modifications similar to those of higher eukaryote organisms; generating a correct folding of the protein (Viader-Salvadó & Guerrero-Olazarán, 2010).

This yeast has become one of the most important expression systems in the production of proteins, because it has a strong promoter and controlled which is the enzyme *alcohol oxidase* (pAOX1). It uses methanol as a source of carbon and energy and posses *histidinol dehydrogenase* gene (his4) for the synthesis of the amino acid histidine (Rivero et al., 2016). Therefore, as a eukaryotic cell *Pichia Pastoris* yeast has many of the advantages of higher eukaryotic expression systems, such as protein processing, protein folding, and post-translational modification which makes it easier to manipulate. It is faster, easier and less expensive to use than other eukaryotic expression systems and has the additional advantage of generating expression levels of heterologous proteins between 10 and 100 times higher. These characteristics make Pichia very useful as a protein expression system (Invitrogen, 2010).

2.2. Pichia Pastoris Characterization

Microscopically are positive large eukaryoticells, with a lenght of 9.6 ($\pm$0.2) Mpb, organized into 4 chromosomes and around 5313 coding genes (Espejo, 2016). However, this strain highlights over other types of host due to its relatively fast growth rate in culture media. The culture is mainly composed of a carbon source, whether glucose, glycerol or methanol, and the last one is the most frequently used (Rivero et al., 2016). Moreover, it requires minimal, simple and economical factors in recombinant protein and cell growth, such as effects of: methanol concentration, dissolved oxygen concentration, induction temperature, pH and nitrogen concentration (Viader-Salvadó & Guerrero-Olazarán, 2010).

On the other hand, expression systems require a method of transferring the DNA sequence of interest to the host cell together with a promoter capable of controlling the production of the foreign genetic product. Successful promoters have very high transcription efficiency, and these are very cheap (Vedvick, 1991), however, most of the foreign protein genes expressed in *P. Pastoris* strains have shown high copy numbers and numbers and higher number of expression cassettes. Thus cassettes. Thus greater amount of protein are produced (Vedvick, 1991).

2.3. Necessary conditions for the recombinant protein production using the P. Pastoris

To obtain high levels of modified ovalbumin in the crop, the strain is required to have the optimal conditions of substrate, dissolved oxygen, temperature, pH and nitrogen that allow the protein to be developed properly.

2.3.1. Effects of methanol concentration

The presence of methanol as an energy source is important for the culture medium medium because the

transcription levels of the heterologous protein depend on the amount of methanol presented by the expression system of the enzyme alcohol oxidase (AOX). If there is shortage or excess of methanol in the culture it would impair the transcriptional efficiency of AOX and it cause the accumulation of formaldehyde by contact of the dissolved oxygen (DO). A possible result the protein expression rate would be affected. Mayson, Kilburn and their co-authors in 2003 suggested that the percentage rate of methanol in the culture medium for heterologous proteins varies from a range of 0.1 to 3.0 % (v/v) , since the methanol feeding system at growth-limiting rates could be 3 to 5 times higher than with excessive methanol feeding (Rivero et al., 2016; Mayson et al., 2003).

When starting the methanol feed, the gene is rapidly and fully transcribed. The strength of the promoter is demonstrated by the observation that the enzyme AOX comprises up to 30% of the soluble protein in extracts of P. pastoris grown in methanol (Vedvick, 1991).

2.3.2. Effects of dissolved oxygen concentration

The use of methanol in the presence of oxygen is the first step in the assimilation of carbon sources, as well as in obtaining energy from it, generating with it the formation of formaldehydes [CH_3OH-CH_2O], a chemical that increases the production of recombinant proteins (GAO & SHI, 2013), for this reason it is important to note that the metabolism of methanol in the presence of large quantities of P. *Pastoris* cultures results in an increase in the demand for DO, so that when cells grow under limiting conditions of DO decreases production levels of the heterologous protein (Rivero et al., 2016; Mayson et al., 2003); for this reason there are several strategies that can be used to maintain the ideal concentration of DO in the culture medium, such as (Rivero et al., 2016):

1. Increasesr airflow.
2. Increasesr the rpm of the stirring process, which generates more oxygen in the culture medium.
3. Cultivar la cepa P. Pastoris at temperatures below 30 °C.
4. Increasers the ratio of the Air/O_2 mixture.
5. A use the ratio of the mixture of O_2/N_2.
6. Controls methanol de feeding.

However, the dissolved oxygen ideonee concentration in the culture medium can range from 20% to 30% (Rivero et al., 2016; GAO & SHI, 2013).

2.3.3. Effets of indution temperature

The effect of temperature on cell growth and the production of recombinant proteins on the endoplasmic reticulum of yeast, can be relevant due to if very high temperatures are used, folding can occur incorrectly, causing the degradation of the same and hence the stress of the strain. This would cause a metabolic overload in it, so the use of optimal temperature is very important. The feeding of the substrate during the growth of the strain and methanol induction phases is 30°C, the metabolic stress of P. *Pastoris* and the formation of toxic products decrease significantly, causing cell growth and the production of recombinant proteins to increase (Rivero et al., 2016). On the other hand, some research has reported that reducing the induction temperature of methanol from 30°C to 20°C is beneficial for the production of recombinant proteins, because at lower induction temperature, activation of alcohol oxidase (AOX) would increase the oxygen uptake rate (OUR), alleviating cell skeleton lysis and the secretion of protease, proteolytically reducing extracellular activity. However, doing this temperature drop and increasing OUR could be an inconvenience in the industry, as this increase would be a problem in terms of energy and as a result economic losses (GAO & SHI, 2013).

Different studies have shown that the growth temperature of Pichia pastoris is 28 - 30°C for liquid cultures, plates, and slants. Growth above 32°C during induction can be detrimental to protein expression and can even lead to cell death (Invitrogen, 2010).

2.3.4. Effects of pH

The *P. Pastoris* strain has a wide pH range to which they can be used, these can range from a pH range of 3.0 to 7.0, however, the optimal pH for strain and methanol growth as an inducer for the production of heterologous proteins is approximately 3.5 and 5.5, depending on the nature of the recombinant protein, for this, the pH has to be adjusted during the growth of the strain to a pH of 5.0 and this has to be adjusted again to a pH of 4.5 to introduce the methanol (Rivero et al., 2016; GAO & SHI, 2013).

2.3.5. Effects of nitrogen concentration

Nitrogen concentration is generally used to control the pH of the crop and for this, Yang and his co-authors in 2004 determined that the production of recombinant huridine by ammonium hydroxide (NH_4OH) was obtained at a concentration of 0.4 M (Rivero et al., 2016; Yang et al., 2004).

3. Methodology

3.1 Determine the DNA sequence coding for the selected protein

The production of the recombinant protein is carried out through the eukaryotic organism *Pichia Pastoris,* due to the advantages it has such as the processing and folding of proteins, in addition to its easy manipulation for it. On the other hand, the system that uses *Pichia Pastoris* as a host is simpler and less expensive than others (Cab-Barrera, 2000). The protein of interest to be manufactured is Ovalbumin, the modified sequence of amino acids for Ovalbumin (Agresott et al., 2020) is expressed below (see Figure 1) and corresponds to the approach developed in the first delivery.

MGSIGAASMEFCFDVFKELKVHHANENIFYCPIAIMSALAMVYLGARQSTRTQINKVVRF
DKLPGFGDSIEAQCGTSVNVHSSLRDILNQITKPNQVYSFSLASRLYAEERYPILPEYLQ
CVKELYRGGLEPINFQTAADQARELINSWVESQTNGIIRNVLQPSSVDSQTAMVLVNAIV
FKGLWEKAFKDEDTDAMYFKVTEQESKPVQMMYQIGLFRVASMASEKMKILELPFASGTM
SMLVLLYQFVSGLEDLESIIQFEKLTEWTSSNVMEERKIKVYLPRMKMEEKYNLTSVLMA
MGITDVFSSSANLSGISSAESLKISQAVHAAHAEINEAGREVVGSAEAGVDAASVSEEFR
ADHPFLFCIKHIATNAVLFFGRCVSP

Figure 1: Sequence of the protein to be cloned (modified Ovalbumin).

3.1.1 Reverse translation

Starting from the modified sequence of amino acids for the Ovalbumin protein (Figure 1) and by means of the online tool *Sequence Manipulation Suite (www.bioinformatics.org)* reversed translation of amino acid sequence was performed. Simultaneously in the same interface of *Sequence Manipulation Suite* you enter the most likely s codon usage table for our *Pichia Pastoris* agency (this due to the snout bias) illustrated below.

AmAcid	Codon	Number	/1000	Fraction ..	AmAcid	Codon	Number	/1000	Fraction ..
G	GGG	468.00	5.76	0.00	W	TGG	834.00	10.26	0.00
G	GGA	1550.00	19.06	0.00	END	TGA	27.00	0.33	0.00
G	GGT	2075.00	25.52	0.00	C	TGT	626.00	7.70	0.00
G	GGC	655.00	8.06	0.00	C	TGC	356.00	4.38	0.00
E	GAG	2360.00	29.03	0.00	END	TAG	40.00	0.49	0.00
E	GAA	3043.00	37.43	0.00	END	TAA	69.00	0.85	0.00
D	GAT	2899.00	35.66	0.00	Y	TAT	1300.00	15.99	0.00
D	GAC	2103.00	25.87	0.00	Y	TAC	1473.00	18.12	0.00
V	GTG	998.00	12.28	0.00	L	TTG	2562.00	31.51	0.00
V	GTA	804.00	9.89	0.00	L	TTA	1265.00	15.56	0.00
V	GTT	2188.00	26.91	0.00	F	TTT	1963.00	24.14	0.00
V	GTC	1210.00	14.88	0.00	F	TTC	1675.00	20.60	0.00
A	GCG	314.00	3.86	0.00	S	TCG	598.00	7.36	0.00
A	GCA	1228.00	15.10	0.00	S	TCA	1234.00	15.18	0.00
A	GCT	2351.00	28.92	0.00	S	TCT	1983.00	24.39	0.00
A	GCC	1348.00	16.58	0.00	S	TCC	1344.00	16.53	0.00
R	AGG	539.00	6.63	0.00	R	CGG	158.00	1.94	0.00
R	AGA	1634.00	20.10	0.00	R	CGA	340.00	4.18	0.00
S	AGT	1020.00	12.55	0.00	R	CGT	564.00	6.94	0.00
S	AGC	621.00	7.64	0.00	R	CGC	175.00	2.15	0.00
K	AAG	2748.00	33.80	0.00	Q	CAG	1323.00	16.27	0.00
K	AAA	2433.00	29.93	0.00	Q	CAA	2069.00	25.45	0.00
N	AAT	2038.00	25.07	0.00	H	CAT	960.00	11.81	0.00
N	AAC	2168.00	26.67	0.00	H	CAC	737.00	9.07	0.00
M	ATG	1517.00	18.66	0.00	L	CTG	1215.00	14.94	0.00
I	ATA	906.00	11.14	0.00	L	CTA	873.00	10.74	0.00
I	ATT	2532.00	31.14	0.00	L	CTT	1289.00	15.85	0.00
I	ATC	1580.00	19.43	0.00	L	CTC	620.00	7.63	0.00
T	ACG	491.00	6.04	0.00	P	CCG	320.00	3.94	0.00
T	ACA	1118.00	13.75	0.00	P	CCA	1540.00	18.94	0.00
T	ACT	1820.00	22.39	0.00	P	CCT	1282.00	15.77	0.00
T	ACC	1175.00	14.45	0.00	P	CCC	553.00	6.80	0.00

Figure 2: Codon usage from Pichia Pastoris (www.kazusa.or.jp).

With this clear, the next step was to obtainer the sequence of AND encoding the sequence of inserted amino acids (enhanced sequence of amino acids for Ovalbumin), that is, the sequence of AND to be cloned without optimization of codons for the Ovalbumin of interest (see Figure reported in results and discussion).

3.2 Optimized sequence of codons for the Pichia Pastoris playback platform

Through the online application *Benchling (www.benchling.com)* develops the project for the cloning of the protein of interest, for this the optimized sequence of codons will be inserted into the *Benchling* tool.

3.2.1. Plasmid selection

To select the plasmid to use first, the vector expression for *Pichia Pastoris* is searched using the *EMBL* online tool (*www.embl.de*). Due to the great use that has commercial mind *pPIC9K* (with promoter AOX1) was selected as a vector to carry out cloning in *Pichia Pastoris* (Invitrogen, 2010). For the download of the *pPIC9K* plasmid map the *SnapGene* software *(www.snapgene.com)*, this map was of great importance because it made the relevant modifications for the cloning of the protein in the *Benchling* tool.

6
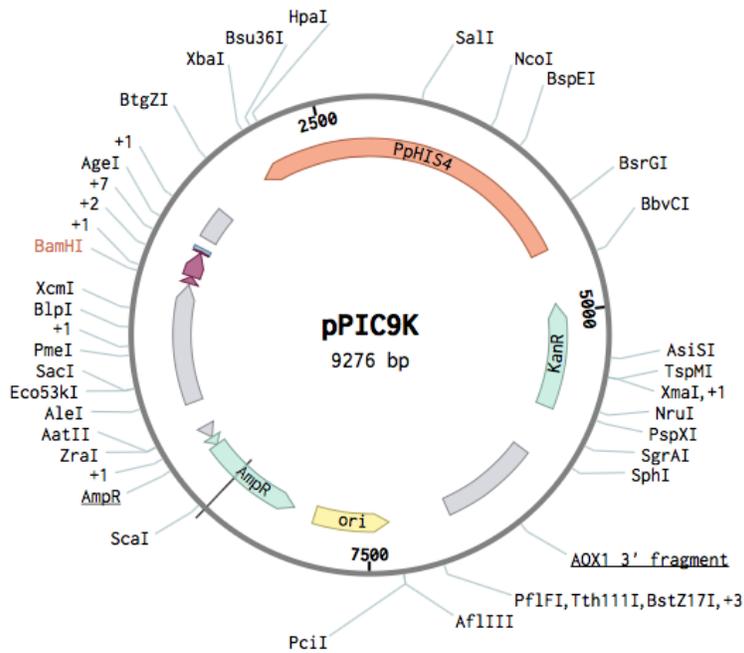
Figure 3: Plasmid Map from pPIC9K (Invitrogen, 2010).

When is open the vector map within the *Benchling* online tool, the insertion of the codon sequence within the plasmid is located in the region followed by the promoter and previous of the MCS. For the selection of restriction enzymes, we work with the cutting enzymes listed by NEB within the *Benchling* tool.

Where it is observed that the plasmid promoter is found the enzyme *BamHI* which makes a cohesive cut, the other restriction enzyme selected is *PsiI*, this is located at the end of the area where the gene is introduced. The information contained between the two plasmid enzymes (*BamHI* and *PsiI*) is replaced to clone there the AND sequence of the modified Ovalbumin protein, the procedure consists of disposition of these enzymes at the beginning and end of the optimized sequence of codons for ovalbumin. When cutting the sequence with *BamHI* and *PsiI*, the entire sequence is taken from *BamHI* to the site where it cuts the *PsiI* enzyme all this based on the leading strand of the AND chain, simultaneously and based on the complementary thread is taken what includes between the enzyme *PsiI* to the cutting of the enzyme *BamHI*, the product of doing this is the sequence which is subsequently inserted into the plasmid *pPIC9K* *with the help of the ligase enzyme.*
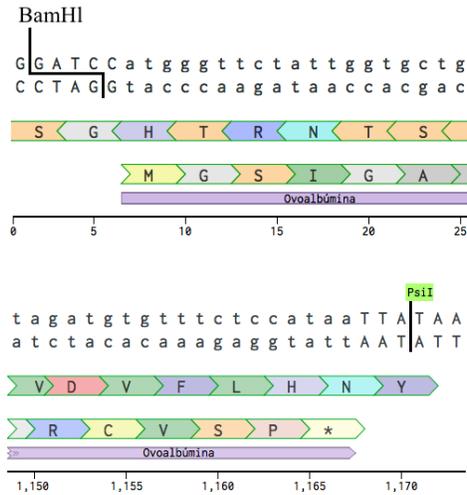
Figure 4: Cut site from enzyme BamHl (recognizes the sequence: GGATCC) this cuts at base 939 of the plasmid leader chain and Psil (recognizes the sequence: TTATAA) that cuts at base 1145 of the plasmid leader chain.

After obtaining the sequence of codons of the ovalbumin protein inserted in the plasmid *pPIC9K* gives way to the realization of optimization for the sequence of codons for which we have the *Benchling* tool, to carry out this process taking into account that the host organism corresponds to *Pichia Pastoris*, it also generates the elimination of sequences of restriction enzymes that are not necessary for the case (the enzymes of restrictions must be protected so as not to be removed), and defining the GC content between 0.33 – 0.66 this in order to have greater stability in the sequence to be cloned. The results are reported in Figure 9.

3.3. Design for gene cloning in pPIC9K plasmid for Pichia Pastoris

After obtaining the optimized la codon sequence for Ovalbumin and through the *Benchiling's* online tool, the step is to perform cloning, this cloning is executed with a vector corresponding to the plasmid *pPIC9K*. For this, the first to consider is to establish the coding gene of the protein (as well as the single cutting sites corresponding to the Enzymes *BamHI* and *PsiI*). The gene to be sequence for the modified ovalbumin corresponds to the optimized sequence of codons subsequently developed (see Figure 9). This is done by considering that what is going to occur corresponds to the optimized sequence of codons plus the restriction site (because the restriction site allows us to "cut" and "paste" our sequence into our *pPIC9K* plasmid).Restriction enzymes cut into DNA, preferably restriction enzymes are not left at the ends of the chain, so it's convenient to additionally take a number of nucleotides before *BamHI* and after *PsiI*, then the sequence to be synthesized (including the optimized sequence of codons for ovalbumin + cutting sites + additional nucleotides) corresponds to the one shown in Figure 5.

additional nucleotides			BamHI		ovalbumine optimized codon sequence	PsiI		additional nucleotides		
TTA	TTC	GAA	GGA	TCC	(Length 1161)	TTA	TAA	ATA	CTA	CTA

Figure 5: Sequence to synthesize.

Subsequently, the DNA fragment found above is used to determine its size, this is done through virtual digestion, where the restriction enzymes *BamHI* and *PsiI* are used.

8

3.3.1. Virtual Digestion

Cutting sites are established by the restriction enzymes *BamHI* and *PsiI*, for this process it should benoted that restriction enzymes are only functional under certain temperature conditions, pH, buffer.

NOTE: The selected enzymes are from *NEB* (*New England Biolabs*), for which the provider reports the following conditions of functionality in Figure 6.

Enzymes	Cuts	Temp.	1.1	2.1	3.1	4/CS
BamHI	1	37°C	75'	100'	100	100'
Psil	1	37°C	10	100'	101	100'

Figure 6: Functionality conditions for restriction enzymes.

Therefore, the original plasmid without any modification must be digested by both restriction enzymes, the virtual digestion obtained for the plasmid *pPIC9K* in its factory state is observed in column 1 of Figure 10.

(a)   (b)
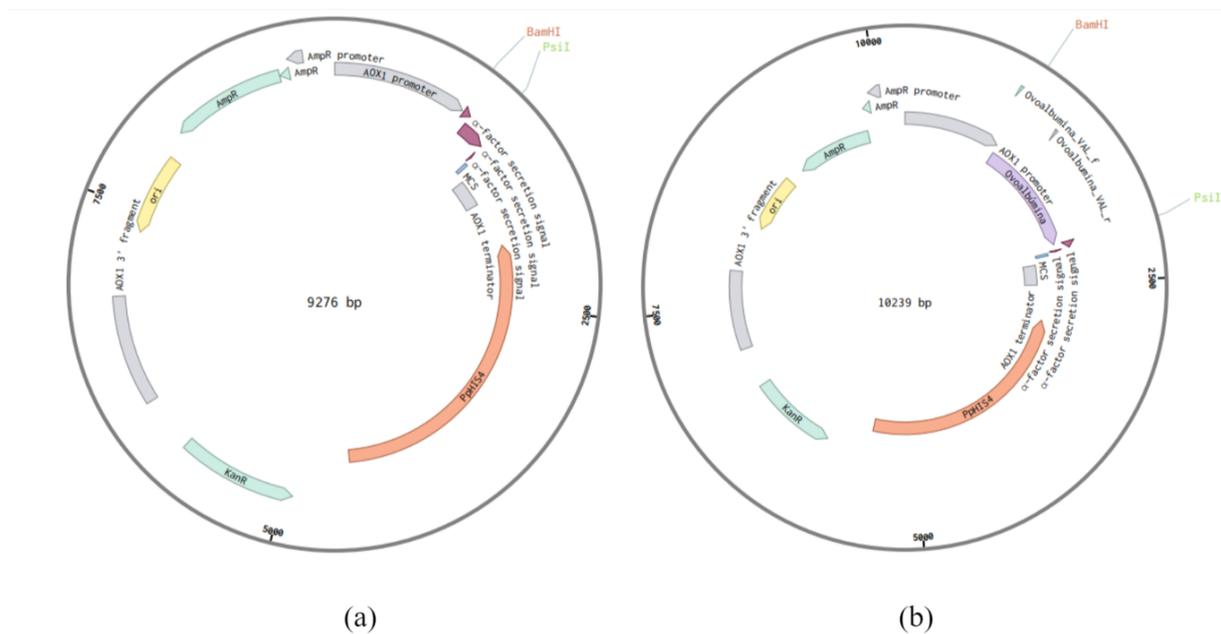
Figure 7: pPIC9K plasmid with restriction enzymes: (a) original and (b) with optimized sequence of codons.

On the other hand, when we digest the plasmid that possesses the optimized sequence of codons for the ovoalbumin protein with the cutting enzymes *BamHI* and *PsiI*, the result reported in column 2 of Figure 10 is obtained.

3.3.2. Linking the gene to be cloned and the plasmid

With the above digestions, the next step is to perform the ligation between the gene to be cloned and the plasmid, to perform this ligation and run it in the same gel we are located in the plasmid that has inserts the optimization of codons and we digest this with the enzyme of restriction *PsiI* that cuts only once (this is done because the plasmid has a circular shape, but when cut with the enzyme the plasmid opens up linear). Reporting in column 3 of Figure 10.

In this way the plasmids to be cloned are built and the insertion of the ovalbumin protein is validated correctly in the plasmid.

3.3.3. Polimerase chain reaction

To perform the PCR process is part of the design of a Primer *in the region before the gene to be cloned and a* Primer located within the same gene (these Primers must have a melting temperature around 58°C and an approximate size of between 18 to 25 base pairs for each). The Primer located in the forward position has 24 base pairs with a melting temperature of 58.00°C and the other Primer in reverse orientation consists of 24 base pairs with a melting temperature of 58.37°C. Subsequently, the two *Primers* are linked to the product to obtain PCR, obtaining 412 pairs of bases that can be seen in Figure 8.
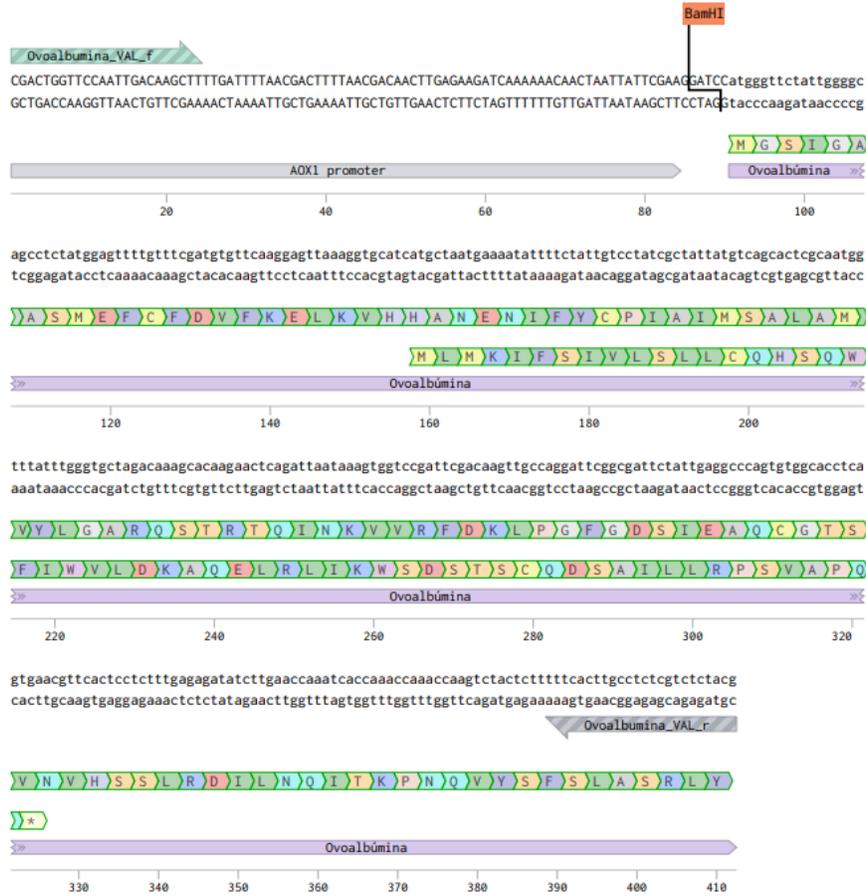
Figure 8: PCR Product

4. Results and Discussion

4.1. DNA sequence of modified ovoalbumina without optimizing its codons and DNA sequence with the optimization of them.

In this figure 9 is the comparison of the DNA sequence of modified ovoalbumin without optimizing its codons and DNA sequence with the optimization of them, where it was found that 196 codons belonging to the amino acid chain of the modified ovoalbumin, which are reflected in gray.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Codon	atg	ggt	tct	att	ggt	gct	gct	tct	atg	gaa	ttt	tgt	ttt	gat	gtt	ttt	aag	gaa	ttg	aag
Codon Optimized	atg	ggt	tct	att	ggg	gca	gcc	tct	atg	gag	ttt	tgt	ttc	gat	gtg	ttc	aag	gag	tta	aag
Aminoacid	M	G	S	I	G	A	A	S	M	E	F	C	F	D	V	F	K	E	L	K

	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Codon	gtt	cat	cat	gct	aac	gaa	aac	att	ttt	tac	tgt	cca	att	gct	att	atg	tct	gct	ttg	gct
Codon Optimized	gtg	cat	cat	gct	aat	gaa	aat	att	ttc	tat	tgt	cct	atc	gct	att	atg	tca	gca	ctc	gca
Aminoacid	V	H	H	A	N	E	N	I	F	Y	C	P	I	A	I	M	S	A	L	A

	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Codon	atg	gtt	tac	ttg	ggt	gct	aga	caa	tct	act	aga	act	caa	att	aac	aag	gtt	gtt	aga	ttt
Codon Optimized	atg	gtt	tat	ttg	ggt	gct	aga	caa	agc	aca	aga	act	cag	att	aat	aaa	gtg	gtc	cga	ttc
Aminoacid	M	V	Y	L	G	A	R	Q	S	T	R	T	Q	I	N	K	V	V	R	F

	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
Codon	gat	aag	ttg	cca	ggt	ttt	ggt	gat	tct	att	gaa	gct	caa	tgt	ggt	act	tct	gtt	aac	gtt
Codon Optimized	gac	aag	ttg	cca	gga	ttc	ggc	gat	tct	att	gag	gcc	cag	tgt	ggc	acc	tca	gtg	aac	gtt
Aminoacid	D	K	L	P	G	F	G	D	S	I	E	A	Q	C	G	T	S	V	N	V

	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Codon	cat	tct	tct	ttg	aga	gat	att	ttg	aac	caa	att	act	aag	cca	aac	caa	gtt	tac	tct	ttt
Codon Optimized	cac	tcc	tct	ttg	aga	gat	atc	ttg	aac	caa	atc	acc	aaa	cca	aac	caa	gtc	tac	tct	ttt
Aminoacid	H	S	S	L	R	D	I	L	N	Q	I	T	K	P	N	Q	V	Y	S	F

	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
Codon	tct	ttg	gct	tct	aga	ttg	tac	gct	gaa	gaa	aga	tac	cca	att	ttg	cca	gaa	tac	ttg	caa
Codon Optimized	tca	ctt	gcc	tct	cgt	ctc	tac	gct	gag	gaa	aga	tac	cct	att	tta	cct	gag	tac	tta	caa
Aminoacid	S	L	A	S	R	L	Y	A	E	R	Y	P	I	L	P	E	Y	L	Q

	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Codon	tgt	gtt	aag	gaa	ttg	tac	aga	ggt	ggt	ttg	gaa	cca	att	aac	ttt	caa	act	gct	gct	gat
Codon Optimized	tgt	gtc	aag	gag	ttg	tat	aga	gga	ggt	tta	gaa	cca	att	aac	ttc	caa	aca	gcg	gct	gac
Aminoacid	C	V	K	E	L	Y	R	G	G	L	E	P	I	N	F	Q	T	A	A	D

	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160
Codon	caa	gct	aga	gaa	ttg	att	aac	tct	tgg	gtt	gaa	tct	caa	act	aac	ggt	att	att	aga	aac
Codon Optimized	cag	gca	cgg	gaa	ctt	att	aac	tcc	tgg	gtt	gag	agt	caa	acc	aac	ggt	atc	atc	cgt	aat
Aminoacid	Q	A	R	E	L	I	N	S	W	V	E	S	Q	T	N	G	I	I	R	N

	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180
Codon	gtt	ttg	caa	cca	tct	tct	gtt	gat	tct	caa	act	gct	atg	gtt	ttg	gtt	aac	gct	att	gtt
Codon Optimized	gta	ttg	caa	cct	tcg	agt	gtt	gat	tcc	cag	act	gcc	atg	gta	ctt	gtg	aat	gct	atc	gtt
Aminoacid	V	L	Q	P	S	S	V	D	S	Q	T	A	M	V	L	V	N	A	I	V

	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200
Codon	ttt	aag	ggt	ttg	tgg	gaa	aag	gct	ttt	aag	gat	gaa	gat	act	gat	gct	atg	tac	ttt	aag
Codon Optimized	ttt	aag	gga	ttg	tgg	gaa	aag	gct	ttt	aag	gac	gaa	gat	aca	gat	gca	atg	tat	ttt	aag
Aminoacid	F	K	G	L	W	E	K	A	F	K	D	E	D	T	D	A	M	Y	F	K

	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220
Codon	gtt	act	gaa	caa	gaa	tct	aag	cca	gtt	caa	atg	atg	tac	caa	att	ggt	ttg	ttt	aga	gtt
Codon Optimized	gtc	act	gaa	caa	gaa	tcc	aaa	ccc	gtt	cag	atg	atg	tac	cag	atc	ggt	ttg	ttc	aga	gtg
Aminoacid	V	T	E	Q	E	S	K	P	V	Q	M	M	Y	Q	I	G	L	F	R	V

	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240
Codon	gct	tct	atg	gct	tct	gaa	aag	atg	aag	att	ggt	gaa	ttg	cca	ttt	gct	tct	ggt	act	atg
Codon Optimized	gcc	agc	atg	gct	tct	gag	aag	atg	aaa	att	ctt	gaa	ttg	cca	ttt	gct	tct	ggg	acg	atg
Aminoacid	A	S	M	A	S	E	K	M	K	I	L	E	L	P	F	A	S	G	T	M

	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260
Codon	tct	atg	ttg	gtt	ttg	ttg	tac	caa	gaa	gtt	tct	ggt	ttg	gaa	gat	ttg	gaa	tct	att	att
Codon Optimized	agt	atg	ctg	gtt	ctc	ttg	tat	cag	gag	gtt	tct	ggt	ctg	gag	gac	cta	gaa	tct	att	att

4.2. Design for gene cloning in pPIC9K plasmid for Pichia Pastoris

When performing virtual digestions and then an agarous gel by means of electrophoresis to determine the size of the bands for the plasmid *pPIC9K* in factory state, *pPIC9K* with optimized sequence of codons (double cut *BamHI* and *PsiI*) and *pPIC9K* with optimized sequence of codons (1 *PsiI* cut only) represented in columns 1, 2, 3 of Figure 10 respectively.
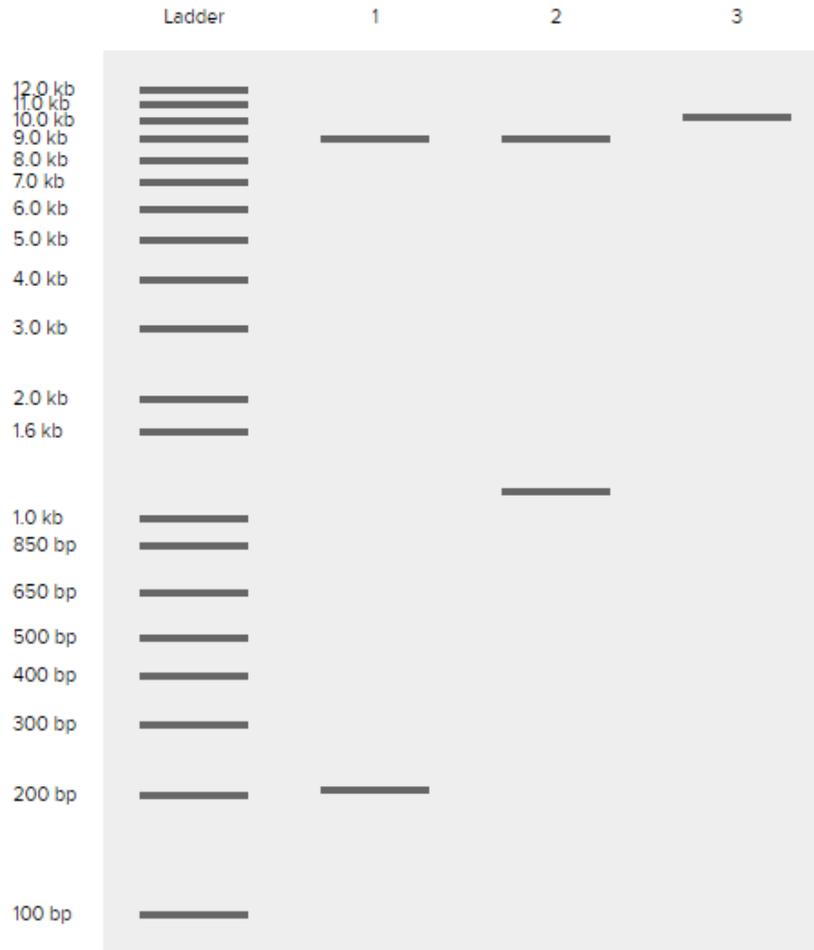
Figure 10: Electrophoresis performed (1: Factory pPIC9K, 2: pPIC9K with codon optimization, 3: pPIC9K with a single cut).

Column 1 of Figure 10 developed the electrophoresis for *pPIC9K* in its original state, where a higher band corresponding to a size of 9.1 kb of the acquired plasmid is observed. On the other hand in column 2 of Figure 10, two fragments are observed, the smaller fragment corresponds to the ovalbumin protein, which is the gene to be cloned with a size of 1.2 kb, while the other fragment corresponds to the plasmid *pPIC9K* of 9.1 kb. By linking the gene to be cloned (lower band column 2) and plasmid (upper band column 1) the digestion observed in column 3 is obtained with a band size of 10.2 kb, which validates the insertion of the ovalbumin protein within the plasmid *pPIC9K*.

5. Conclusions

The sequence of codons optimized using the computational tools allows to guarantee the stabilization of the sequence of this gene for the execution of its cloning in *Pichia Pastoris* with *pPIC9K* as a cloning vector (under the functionality conditions of the *BamHI* and *PSiI* restriction enzymes). In addition, through electrophoresis it is possible to validate the correct insertion of the optimized sequence to be cloned within the plasmid, for its subsequent downstream production strategy.

References

Milk and Dairy Products as Functional Foods: A Review. (2011). *International Journal of Dairy Science*, *6*(1), 1–12. https://doi.org/10.3923/ijds.2011.1.12

Epidemiology of Food Allergy in Latin America. (2015). *Allergol Immunopathol, 43*.

"*Guidance on Food Allergen Management for Food Manufacturers*". (2013).

Cosmetic Science and Technology. (2009). In *Oil Spill Science and Technology*. Elsevier. https://doi.org/10.1016/b978-1-85617-943-0.10041-3

A review: Chemical Composition and Utilization of Egg. (2018). *Int. J. Chem. Stud.*, *6, 3186–3189*.

El Gran Libro del Huevo. (2009). Editorial Everest S.A. http://institutohuevo.com/wp-content/uploads/2017/07/EL-GRAN-LIBRO-DEL-HUEVO.pdf

Fine Mapping and Structural Analysis of Immunodominant IgE Allergenic Epitopes in Chicken Egg Ovalbumin. (2003). *Protein Engineering Design and Selection*, *16*(10), 747–752. https://doi.org/10.1093/protein/gzg095

Structure and Properties of Ovalbumin. (2001). *Journal of Chromatography B: Biomedical Sciences and Applications*, *756*(1-2), 189–198. https://doi.org/10.1016/s0378-4347(01)00108-6

BIOLOGIA CELULAR Y MOLECULAR: Conceptos y Experimentos. (2011). McGraw-Hill Interamericana de España.

Biotechnology. (2015). Elsevier Science. https://books.google.com/books/about/Biotechnology.html?hl=es&id=osqcBAAAQBAJ

Insoluble Proteins: Methods and Protocols. (2014). Humans Press. http://bio1.ir/images/pdf/96/10/book/Insoluble%20Proteins_%20Methods%20and%20Protocols-Springer%20New%20York%20(2015).pdf

Galindo, E., & Ramírez, O. T. (Eds.). (1998). *Advances in Bioprocess Engineering*. Springer Netherlands. https://doi.org/10.1007/978-94-017-0643-8

Pichia Pastoris: Una Plataforma Para la Producción de Proteínas Heterólogas. (2016). *Revista CENIC. Ciencias Biológicas*, *47, 67–77*.

Comparación Teórica de las Capacidades Metabólicas de Saccharomyces Cerevisiae y Pichia Pastoris para la Producción de SOD. (2007). Universidad de Chile Facultad de Ciencias Físicas y Matemáticas Departamento de Ingeniería Química y Biotecnología. http://repositorio.uchile.cl/tesis/uchile/2007/cominetti_oa/sources/cominetti_oa.pdf

Generation of Diploid Pichia Pastoris Strains by Mating and Their Application for Recombinant Protein Production. (2012). *Microbial Cell Factories*, *11*(1), 91. https://doi.org/10.1186/1475-2859-11-91

Biotecnología de Proteínas Recombinantes con Pichia Pastoris. (2010). *XIV Congreso Nacional De Biotecnología y Bioingeniería*. https://smbb.mx/congresos%20smbb/queretaro11/TRABAJOS/simposios/SimposioVI_Viader.pdf

Pichia Expression Kit for Expression of Recombinant Proteins in Pichia Pastoris. (2010). Life Technologies. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/pich_man.pdf

Producción, Purificación, Caracterización de Tres Hexosaminidasas Lisosomales Recombinantes en Pichia Pastoris. (2016). Pontifica Universidad Javeriana. 36-37. https://repository.javeriana.edu.co/bitstream/handle/10554/19644/EspejoMojicaAngelaJohana2016.pdf?sequence=1&isAllowed=y

Gene expression in yeast: Pichia Pastoris. (1991). *Current Opinion in Biotechnology*, *2*(5), 742–745. https://doi.org/10.1016/0958-1669(91)90045-7

Effects of Methanol Concentration On Expression Levels of Recombinant Protein in Fed-batch Cultures of Pichia Methanolica. (2003). *Biotechnology and Bioengineering*, *81*(3), 291–298. https://doi.org/10.1002/bit.10464

Process Control and Optimization for Heterologous Protein Production by Methylotrophic Pichia pastoris. (2013). *Chinese Journal of Chemical Engineering*, *21*(2), 216–226. https://doi.org/10.1016/s1004-9541(13)60461-9

Improvement of Recombinant Hirudin Production by Controlling NH4+ Concentration in Pichia Pastoris Fermentation. (2004). *Biotechnology Letters*, *26*(12), 1013–1017. https://doi.org/10.1023/b:bile.0000030049.75092.95

Construccion de una cepa de Pichia Pastorís Sobreproductora de la Isoforma de 20 Kda de la Hormona Del Crecimiento Humano. (2000). [Master's thesis, http://eprints.uanl.mx/6421/1/1080111679.PDF]. http://eprints.uanl.mx/6421/1/1080111679.PDF

Allergen Modification Of The Ovalbumin As Possible Method To Minimize The Egg Allergenic Reaction Into The Human Organism And Its Inclusion in a Food Product. (2020). *Autherea*. https://www.authorea.com/users/308685/articles/439669-allergen-modification-of-the-ovalbumin-as-possible-method-to-minimize-the-egg-allergenic-reaction-into-the-human-organism-and-its-inclusion-in-a-food-product