Vol. 20, No. 3(2021) Bio2419
Revista Mexicana de Ingeniería Química

Expression of a synthetic protein with a high proportion of essential amino acids by *Pichia pastoris*

Expresión de una proteína sintética con alta proporción de aminoácidos esenciales por *Pichia pastoris*

Y. Bartolo-Aguilar[1], C. Chávez-Cabrera[2,3*], J.C. Cancino-Díaz[3], R. Marsch[1]

[1]*Department of Biotechnology and Bioengineering, Cinvestav-IPN, Av. Instituto Politécnico Nacional 2508, Col. San Pedro Zacatenco, Mexico City, 07360, Mexico.*
[2]*College of Science and Technology Studies of the State of Michoacán, CECyTE Michoacán, Plantel Churumuco, Héroes de la Revolución S/N, Col. Centro, Churumuco de Morelos, Michoacán. 61880, México.*
[3]*Departament of Microbiology, Escuela Nacional de Ciencias Biológicas del Instituto Politécnico Nacional, Prol. Carpio y Plan de Ayala S/N, Col. Santo Tomás, Mexico City. 11340, Mexico.*

Received: April 5, 2021; Accepted: July 2, 2021

Abstract

It is widely known that an adequate intake of proteins with essential amino acids stimulates the production of essential nonprotein substances for the body, such as serotonin, thyroid hormones, and glutathione. Biotechnology applications can focus on food and its benefits for human health, either by improving the nutritional value of existing foods or developing new alternatives that can help resolve the effects of poor nutrition. The yeast *Pichia pastoris* has been designated as a GRAS (generally recognized as safe) microorganism and is currently used to produce a large number of recombinant proteins. In this study, a 100% synthetic protein was expressed intracellularly as a protein with high nutritional quality that contains a 75% ratio of essential amino acids using the expression system of the yeast *P. pastoris*. This approach might offer future possibilities for using the modified yeast as a single-cell protein source in the livestock industry and the food supplement market to reduce total protein intake.

Keywords: essential amino acids, synthetic peptide, human nutrition, *Pichia pastoris*.

Resumen

Es ampliamente conocido que una ingesta adecuada de proteínas con aminoácidos esenciales estimula la producción de sustancias no proteicas esenciales para el organismo, como la serotonina, las hormonas tiroideas y el glutatión. Las aplicaciones de la biotecnología pueden centrarse en los alimentos y sus beneficios para la salud humana, ya sea mejorando el valor nutricional de los alimentos existentes o desarrollando nuevas alternativas que puedan ayudar a resolver los efectos de una mala nutrición. La levadura *Pichia pastoris* ha sido designada como microorganismo GRAS (generalmente reconocido como seguro) y actualmente se usa para producir una gran cantidad de proteínas recombinantes. En este estudio se expresó intracelularmente una proteína 100% sintética como una proteína de alta calidad nutricional que contiene una proporción del 75% de aminoácidos esenciales utilizando el sistema de expresión de la levadura *P. pastoris*. Este enfoque podría ofrecer posibilidades futuras para el uso de la levadura modificada como fuente de proteína unicelular en la industria ganadera y el mercado de complementos alimenticios para reducir la ingesta total de proteínas.

Palabras clave: aminoácidos esenciales, péptido sintético, nutrición humana, *Pichia pastoris*.

* *Corresponding author. E-mail*: a0006102@cecytem.edu.mx
https://doi.org/10.24275/rmiq/Bio2419
ISSN:1665-2738, *issn-e*: 2395-8472

1 Introduction

The WHO (World Health Organization) and the FDA (Food and Drug Administration Organization) report that the number of people in the world affected by malnutrition or chronic lack of food has increased from approximately 804 million in 2016 to almost 821 million in 2017. Extreme climatic events have short-, medium- and long-term repercussions on food security and nutrition, reducing access and availability of quality food (FAO *et al*., 2018). The lack of quality proteins of animal and plant origin in certain regions has become an increasingly common problem (Schulte-Herbrüggen *et al*., 2017); therefore, it is necessary to use alternative sources that satisfy the daily requirements of essential amino acids in people. The intake of proteins with essential amino acids stimulates the production of essential nonprotein substances for the body, such as serotonin, thyroid hormones, and glutathione (Wu, 2009).

Genetic engineering, also called recombinant DNA methodology, is a set of tools and methods that allows the manipulation of genetic material from living organisms. In the mid-1970s, the development and sophistication of recombinant DNA techniques were prompted for two main reasons. The first was to advance knowledge about how genes are organized and regulated in the genome of a living organism. The second was achieving confidence in isolating, altering, and transplanting genes from one organism to another, thus passing on new genetic characteristics to the host organism. However, as with many newly developing fields and technologies, there are risks and fears associated with each possible benefit the new technology brings. Nevertheless, according to Rheeder (2014), there are now more advantages than disadvantages to the further study and development of synthetic biology. The organism that receives genetic material from another organism is called transgenic; in many cases, this new DNA is capable of producing proteins called heterologous proteins. Currently, a wide variety of heterologous proteins, also called recombinant proteins, have been produced, ranging from those of pharmaceutical interest, such as insulin and human growth hormone, to industrial interest, such as pectinases and invertases (Gomes *et al*., 2016; Karbalaei *et al*., 2020), among others. Our working group has studied recombinant enzymes for the investigation of metabolic pathways (Chávez-Cabrera *et al*., 2015) to yeast autolysis by regulation of temperature promoters (Bartolo-Aguilar *et al*., 2017). The most widely used transgenic organisms are bacteria such as *Escherichia coli*, yeasts such as *Saccharomyces cerevisiae* and *Pichia pastoris*, and some higher organisms such as *Arabidopsis thaliana* (Baeshen *et al*., 2014; Çelik and Çalik, 2012) and mammalian cells (Porro *et al*., 2011).

For several decades, there has been interest in obtaining single-cell protein (SCP) for human consumption (Ritala *et al*., 2017). A wide range of fungi have been considered for use as SCPs, and fungi grown as SCPs will generally contain 30-50% protein. The yeasts used for the production of SCP are *S. cerevisiae, Kluyveromyces marxianus* and *P. pastoris*, among others. Methylotrophic yeasts, such as *P. pastoris*, produce biomass and protein from methanol as their source of carbon (Rashad *et al*., 1990). Industrial scale production has been carried out, e.g., by Phillips Petroleum Company. Their yeast produced 130 g L^{-1} biomass, with a productivity of more than 10 g L^{-1} h^{-1} (Johnson, 2013). In addition, *P. pastoris* methylotrophic yeast has been designated by the Food and Drug Administration (FDA) as GRAS (generally recognized as safe) for the production of recombinant proteins (Ciofalo *et al*., 2006; Kim *et al*., 2015) and recombinant biopharmaceuticals (Thompson, 2010; Martínez-Hernández *et al*., 2019). The GRAS status of the FDA has allowed both the single-cell protein and a wide range of protein products to enter the market for therapeutic and industrial uses.

Since 1993, *P. pastoris* dried yeast may be used in feed formulations of broiler chickens as a source of protein not to exceed 10 percent by weight of the total formulation (CFR - Code of Federal Regulations Title 21, 1993). In today's food industry, Impossible Foods' recombinant soy leghemoglobin preparation is another GRAS product and is used to optimize flavor in ground beef analogue products intended to be cooked not to exceed 0.8 percent by weight of the total formulation (Keefe, 2018).

Pichia pastoris technology from RCT (Research Corporation Technologies) reports that there are 16 products produced by *P. pastoris* on the market (https://pichia.com/science-center/commercialized-products/).

Current knowledge of molecular biology, genetic engineering and technological development in the context of the synthesis of highly specific nucleotides provides an opportunity to put forward into research the development of novel peptides that do not exist in nature and are of high nutritional value (Chakraborty *et al*., 2000; Falco *et al*., 1994; Yang *et al*., 1989). Our

groups obtained a 350 bp synthetic gene that encodes a peptide with a high content of essential amino acids by overlapping PCR, which was expressed as a synthetic polypeptide in *E. coli* but was weakly identified by Western blot (Sánchez-Crisóstomo *et al*., 2019). The high metabolic and nutritional demand of the cell culture promoted the low expression of the synthetic protein. These encounters suggested that a 100% synthetic protein with a high content of essential amino acids could enhance its biosynthesis in other expression systems. Additionally, the codon harmonization of a native gene in the host promotes recombinant protein expression by decreasing their transcription speed, among other factors (Angov *et al*., 2008). Thus, the synthetic protein expression with high nutritional value in different expression systems could be feasible.

Therefore, the purpose of this research is that from a synthetic gene designed with codon harmonization, it expresses a 100% synthetic protein of 173 amino acids (SP-6His) and high nutritional value with a high content of essential amino acids, using the yeast *P. pastoris* expression system.

2 Materials and methods

2.1 Strains and culture media

Escherichia coli TOP10 cells were used as a host to preserve and propagate plasmid DNA pPICZB-*SP*, which was purchased from GenScript® Company (New Jersey, USA). *Escherichia coli* TOP10 was grown in Luria-Bertani (LB) medium supplemented with zeocin (30 μg/mL). The yeast *Pichia pastoris* X-33 was used to express the synthetic protein. Recombinant yeasts were grown in 32 mL of minimal medium with 0.5% methanol (MM) in 250 mL baffled flasks (Sigma, MO, USA) according to the manufacturer's directions for the EasySelect™ *Pichia* expression kit (Invitrogen, CA, USA) and incubated for 96 hours on an orbital shaker (New Brunswick Scientific, NJ, USA) at 28°C and 300 rpm. The reagents used for the molecular biology steps were from J. T. Baker, Invitrogen or Sigma.

2.2 Amino acids sequence design (SP)

The amino acid sequence of the synthetic protein is based on the incorporation of essential amino acids to satisfy different metabolic needs, both human (FAO *et al*., 2018) and animal (Kidd and Tillman, 2016), with the additional incorporation of some non-essential amino acids (see Additional file 1: A1). The amino acid sequence organization with a uniform distribution of both acidic and basic amino acids maintains adequate enzymatic digestibility by *in-silico* analysis. Finally, the sequence was verified to have no nuclear localization sequence and no known allergenic epitope (see software in section 2.4).

2.3 Nucleic acids

The design of the synthetic gene sequence reduces the strong mRNA transcription secondary structures to generate the energetically favorable mRNA-*SP* secondary structure for protein expression. The synthetic nucleotide sequence synthesized from the amino acid sequence (SP) considers yeast codons use in combination, both the maximum frequency codons and the intermediate frequency codons for the same amino acid (see Additional file 1: A2). Next, the *SP* gene was synthesized and cloned into the expression vector pPICZB by GenScript® Company (New Jersey, USA). The pPICZB-*SP* construct was preserved in *E. coli* TOP10 and transformed by thermal shock. The pPICZB-*SP* in an *E. coli* TOP10 clone was propagated on LB with zeocin (30 μg/mL) to extract 10 μg of plasmid. The pPICZB-*SP* construct encodes the synthetic protein with a histidine epitope (SP-6His) that allows purification by immobilized metal affinity chromatography (IMAC) (see figure 2 and supplementary material: Additional file 1).

The synthetic gene plus the nucleotide sequence encoding the histidine epitope corresponding to the plasmid pPICZB (hereafter, *SP-6His*) was released with the endonucleases *Eco*RI and *Bam*HI from the plasmid pPICZB-*SP*. The sequence *SP-6His* was cloned into pQE30 to generate the pQE30-*SP* construct (see figure 2). Furthermore, a clone with the construct pQE30-*SP* in *E. coli* BL21 pLyss was also obtained for its intracellular expression of the synthetic protein plus a histidine tag (as a molecular weight control). Plasmids were purified following the Birnboim and Doly protocol (Sambrook and Russel, 2001). The extraction and purification of the plasmids were carried out on a small and medium scale, depending on their application.

2.4 Bioinformatics software

The synthetic SP gene was designed using the Gene Construction Kit™ version 3.5 software, Textco, Inc.,

2011, for the analysis of DNA sequences and to establish the work strategy (supplementary material: Additional file 1) and workflow to construct plasmids expressing the synthetic protein (see figure 2). The analysis of the secondary structures of the mRNA was performed using the RNA fold web server online software. The theoretical digestibility of the peptide was analyzed using Expasy's Peptide Mass software (https://web.expasy.org/peptide_mass/). The essential amino acid profile was determined using Expasy's ProtParam online software (https://web.expasy.org/protparam/). The sequence of the peptide was analyzed with the Structural Database of Allergenic Proteins (http://fermi.utmb.edu/SDAP/) to avoid sequences that correspond to any allergenic epitope in this database.

The spectra generated by the mass spectrometer and corresponding to the peptides of the synthetic protein were identified by the Discovery Proteome software. The relevant parameters include identifying peptides that have up to two trypsin cleavage sites and differ in theoretical molecular weight in 0.6 Da with the monoisotopic m/z. Additional file 1 and figure 2 have a mirror file, each in the Gene Construction kit format (Additional file 3 and additional file 4, respectively).

2.5 *Transformation of the P. pastoris strain*

Pichia pastoris X-33 electrocompetent cells were prepared and electroporated according to the protocol described by Cregg (Cregg, 2007). Identification of transformed yeasts was performed by colony PCR using oligonucleotides *AOX1*-F and *AOX1*-R, according to the protocol of the EasySelect™ *Pichia* expression kit (Invitrogen, CA, USA).

2.6 *Synthetic protein expression*

The expression culture of a pPICZB-*SP* in *P. pastoris* colony was performed according to the EasySelect™ *Pichia* expression kit protocol. The expression of SP-6His with transformed yeast was into two stages: the first stage was producing biomass in YPD culture medium at 28°C and 160 rpm for 16 hours. The second stage was the induction of expression of SP-6His in a minimal medium with 0.5% methanol and with 160 μL methanol for reinduction every 24 hours for 96 hours, at 28°C and 300 rpm.

The expression culture of a pQE30-*SP* in *E. coli* colony was performed according to the QIAexpressionist™ kit protocol, with induction of 1 mM IPTG at 30°C and 160 rpm for 5 h, evaluating protein expression by SDS-PAGE 12% gel.

2.7 *Purification of SP-6His*

The recombinant SP-6His protein was purified using nickel affinity chromatography with Invitrogen™ Ni-NTA Agarose (Invitrogen, CA, USA). Extraction of the soluble protein was performed from the pellet of an expression culture (32 mL) as described by the protocol of Chávez-Cabrera *et al*. (2015).

The soluble extract was mixed with native binding buffer containing 20 mM imidazole (ratio, 1:5 [v/v]), and Invitrogen™ Ni-NTA Agarose (3 mL) was placed in contact with the suspension for 30 minutes and loaded on the column. SP-6His was eluted in four stages by using native elution buffer and increasing concentrations of imidazole (60 mM [25 mL], 80 mM [6 mL], 100 mM [6 mL] and 160 mM [6 mL]). The volumes obtained from the last three stages were concentrated to 500 μL using an Amicon Ultra2 Centrifugal Filter Unit with an Ultracel-50 membrane, 2 mL sample volume (Millipore, Munich, Germany). Finally, the protein profile was analyzed by SDS-PAGE on a 15% gel.

2.8 *Western blot and mass spectrometry (MS) analysis*

The soluble extract containing SP-6His was analyzed by 12% SDS-PAGE, and Western blot analysis was performed using the QIAexpress Detection System (Qiagen, CA, USA) following the manufacturer's instructions. The SP-6His was purified to homogeneity by nickel affinity chromatography, and the band with higher intensity (approx. 25 kDa) was cleaved, digested, alkylated, and desalted according to the Trypsin Profile IGD Kit (Sigma-Aldrich, USA); the mass spectra were analyzed by mass spectrometry (nHPLC-MS, Thermo Scientific), which has an ion trap mass analyzer (orbitrap).

Table 1. *In-silico* digestibility analysis of proteins of high nutritional value.

Enzyme digestive	**SP-This work***	**ovalbumin**	**casein**
	Peptides with the greatest amount of amino acid residues		
Trypsin	20	35	42
Chymotrypsin	17	37	31
Trypsin/Chymotrypsin	18	27	21
Pepsin	14	11	9

*SP: synthetic protein; ovalbumin is ovalbumin isoform X1 (XP_015137652.1); casein is alpha-s1-casein (P02662.2).

3 Results and discussion

3.1 In-silico design of the SP gene

In accordance with the daily requirements considered for an adult as claimed by the World Health Organization (WHO) and the Food and Agriculture Organization of the United Nations (FAO), a protein with high digestibility that meets the essential amino acids (FAO, 2013; Institute of Medicine, 2005) was considered.

Supplementary material: Additional file 1 shows the proposed amino acid sequence as a synthetic protein plus a histidine tag (SP-6His, Additional file 1: A1 and A2) with a high content of essential amino acids. The synthetic peptide shows some acidic amino acids considered nonessential that decrease the hydrophobic load. The high nutritional value synthetic peptide sequence was designed using Gene Construction Kit™ version 3.5 software, Textco, Inc., 2011. The amino acid sequence of SP-6His has 173 amino acids with 129 essential amino acids: 12 His (H), 15 Ile (I), 2 Leu (L), 20 Lys (K), 6 Met (M), 13 Phe (F), 11 Thr (T), 2 Trp (W), and 48 Val (V). SP-6His is a proof of concept for the production of a moderately long (173 amino acids) 100% synthetic peptide in the expression system of yeast *P. pastoris*. The amino acid sequence was organized in order to optimize the digestibility of the peptide with some of the digestive enzymes such as trypsin, chymotrypsin, and pepsin; obtaining amino acid residues with a maximum length of 20, 17, and 14 respectively (see Table 1). The incorporation of 15 negatively charged amino acids, 13 glutamic acids, and 2 aspartic acids are to decrease the positive charge of SP-6His corresponding to the 20 lysine. Valine was incorporated in high proportion into SP-6His to verify the expression of branched-chain amino acids (BCAAs) in the yeast expression system. However, animal studies show that BCAAs are linked to prevent high-fat diet-induced obesity (Newgard *et al*., 2009). In contrast, others report better uptake of fatty acids for lipid build-up (Cummings *et al*., 2017).

On the other hand, knowledge about the production and synthesis capacity of BCAA amino acids in yeasts is minimal, which indicates regulation by the final product. Bussey and Umbarger (1969) suggested that the isoleucine-valine pathway has multivalent and end-product repression in yeast, but changes in enzyme levels are less than those found in bacteria.

Analysis of the protein parameters shows that the isoelectric point is 9.12. A slightly positive hydrophobic charge is 0.58 in the Kyte-Doolittle hydrophobicity scale according to the ProtParam tool from Expasy.

The gene synthesized from the designed amino acid sequence to synthetic protein to correspond for the expression in *Pichia* (see vector pPICZB-*SP* in figure 2 and additional file 1) is 450 bp plus two bp (CC) for the sequence stays in-phase and 67 bp incorporated by the vector pPICZB (522 bp), including three upstream nucleotides (GCC) to make the translation efficient, according to the protocol of the EasySelect™ *Pichia* expression kit (Invitrogen, CA, USA) and the stop codon.

Figure 1 shows the secondary structure of the synthetic mRNA generated. In the case of the RNA model, most of the base pairings are of the Watson-Crick structure. Single-stranded RNA can also form many secondary structures in which a single RNA molecule folds and shapes hairpin loops stabilized by intramolecular hydrogen bonds between complementary bases (Kozak, 2005). Hofacker *et al*. (2004) exploited the RNA structural complexity to generate new RNA species (not found in nature) that have specific desirable properties.

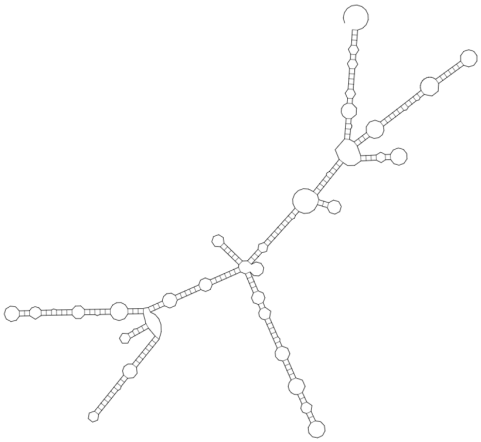

Figure 1. Secondary structure of synthetic mRNA.
The synthetic mRNA encoding the protein with a high content of essential amino acids has a secondary structure with eight hairpin loops, 19 small internal loops, four multibranched loops, and nine helices, which make an unstable secondary structure with a free energy of thermodynamic assembly of -164.92 kcal/mol, according to the RNA fold web server online software.

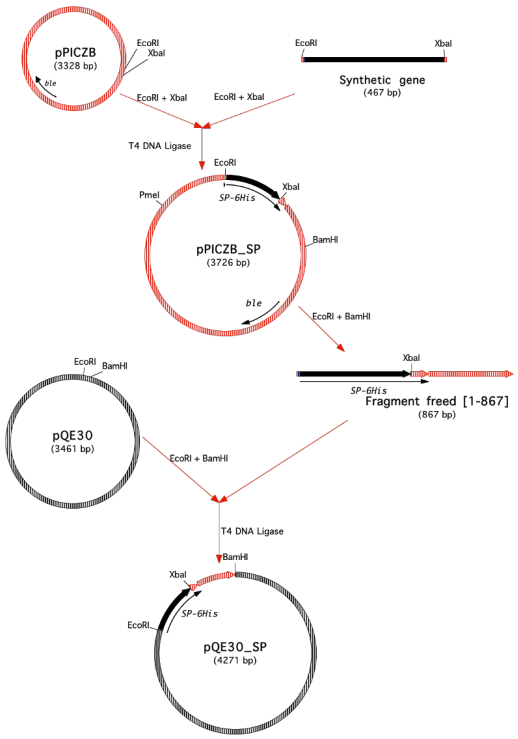

Figure 2. Workflow for construction of plasmid pQE30-*SP* from pPICZB-*SP*.
The strategy for the generation of the new genetic structure was through the Gene Construction Kit Software.

Table 2. Amino acid profiles of high nutritional quality proteins.

Amino acid (AA)	SP-This work % mol	**Ovalbumin** % mol	**Casein** % mol
A	0.6	9	5.6
R	0	3.6	2.8
N	6.4	4.4	3.7
D	1.2	3.6	3.3
C	1.8	1.7	0.5
Q	0.6	3.6	6.5
E	7.6	8.3	11.7
G	0	4.6	4.2
H	7	1.9	2.3
I	8.8	6.3	5.6
L	1.2	8.8	10.3
K	11.7	4.9	7
M	3.5	4.4	2.8
F	7.6	5.8	3.7
P	4.7	3.6	7.9
S	1.8	9.5	7.5
T	6.4	4.4	2.8
W	1.2	0.7	0.9
Y	0	2.7	4.7
V	28.1	8	6.1
%	**75.5**	**45.2**	**47.1**

* AA, amino acid; SP, synthetic protein; ovalbumin is ovalbumin isoform X1 (XP_015137652.1); casein is alpha-s1-casein (P02662.2). Amino acid profiles obtained by Expasy's online ProtParam Tool software. Amino acids marked in red are essential for humans.

Synthetic mRNA from the *SP* gene (mRNA-*SP*) exhibits a minor nucleotide base mismatch secondary structure, as it has eight hairpin loops, 19 small internal loops, four multi-branched loops, and nine helices. Synthetic mRNA-*SP* analysis indicates that the secondary structure has a thermodynamic assembly free energy of -164.92 kcal/mol (Wu *et al.*, 2009; Wu *et al.*, 2018).

According to the RNA fold web server online software, synthetic mRNA-*SP* displays an unstable secondary structure that could prevent early translation interruption with the possibility of correct expression in different expression systems (Mathews *et al.*, 2004). It should be noted that the codons used for the design of the *SP* gene had to be relocated on several occasions to avoid strong secondary structures and obtain free energy of thermodynamic assembly favorable for translation.

Table 2 shows the amino acid profiles of proteins of high nutritional value and, in the column marked "SP-this work," the run-down of the sequence of

synthetic peptides. The designed synthetic protein is 75.5% essential amino acids, while the proteins of high nutritional value are between 45.1% (casein) and 45.2% (ovalbumin).

The sites of action of digestive enzymes were identified with Expasy's Peptide Mass software, the amino acid sequence was combined to achieve high digestion with digestive enzymes; the largest generated peptides were 14 to 20 amino acid residues (see Table 1); and the possible allergenic risks in the proposed peptide were eliminated by the "Structural Database of Allergenic Proteins" (http://fermi.utmb.edu/SDAP/sdap_fas.html) (Ivanciuc *et al*., 2003).

3.2 Synthetic protein expression

The *Pichia* yeast genus is used for the production of aromatic compounds (Renteria-Martínez *et al*., 2021) until the production of recombinant proteins (Martínez-Hernández *et al*., 2019). The choice of yeast *P. pastoris* as an expression system was due to its GRAS status and its current application in producing a unicellular protein with low-cost culture media. These are the main advantages over the expression systems of *E. coli* and mammalian cells. *E. coli* can accumulate endotoxins, while mammalian cell cultures are expensive and prone to contamination with some viruses (Gomes *et al*., 2016).

The *SP* gene encoding the synthetic protein (SP) was incorporated into an expression vector to use the *P. pastoris* intracellular expression system. The sequence of the *SP* gene incorporated into the vector pPICZB was synthesized and sequenced by GenScript® Company (New Jersey, USA).

Unlike the SP-6His protein expressed by a pPICZB-*SP* in the *P. pastoris* colony, the SP-6His protein expressed by the *E. coli* BL21 expression system used as a positive control for molecular protein weight.

Figure 3 shows the analysis of the expression of the synthetic proteins expressed in both expression systems and the preliminary identification by Western blot. Because the soluble protein extract was not quantified, it decided to load decreasing volumes of the protein suspension onto the SDS-PAGE gel to observe immunodetection. Figure 3A shows the protein profile expressed by both expression systems, but overexpression to identify them was not evident. Based on the analysis of the designed amino acid sequence, the molecular weight of the SP-6His protein is 20 kDa, according to the ProtParam tool from

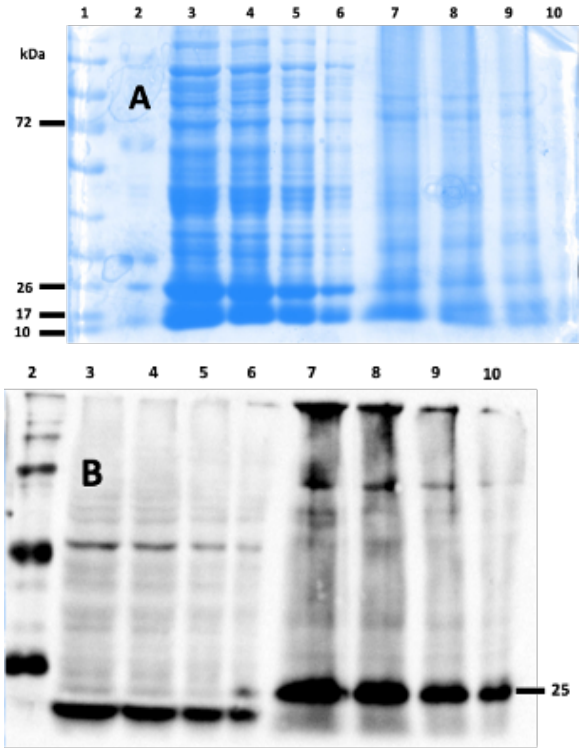

Figure 3. Protein profile and Western blot of the synthetic protein.

Figure 3A: Protein profiles from both expression systems (*E. coli* and *P. pastoris*) and figure 3B: Western blot to identify the synthetic proteins from the respective protein profiles. Figure 3A is shown in Lane 1, molecular weight marker Blue Prestained Protein Standard, Broad Range, BioLabs™. Both figures (3A and 3B) are shown in lane 2, immunodetection positive control (GFP-6His tagged: 27 kDa); lanes 3-6, soluble extract of SP-6His from pQE30-*SP* in *E. coli* (30 μL, 20 μL, 10 μL and 5 μL, respectively) induced with 1 mM IPTG; lanes 7-10, soluble extract of SP-6His from pPICZB-*SP* in *P. pastoris* (30 μL, 20 μL, 10 μL and 5 μL, respectively) induced with 0.5% methanol. The molecular weight of the synthetic protein (SP) *in silico* is 20 kDa. 12% SDS-PAGE gel.

Expasy. Figure 3B shows the protein immunodetection profile in both expression systems and a positive immunodetection control of GFP-6His tagged (27 kDa). Immunodetection with a molecular weight of 20 kDa (Figure 3B: lines 3-6) is shown in the lanes with SP-6His from *E. coli*, while immunodetection with a molecular weight of 25 kDa (Figure 3B: lines 7-10) is shown in the lanes with SP-6His from *P. pastoris*. *Escherichia coli* does not incorporate posttranslational modifications with a high impact on the molecular weight into its expressed recombinant proteins. The

immunodetection a protein of 20 kDa molecular weight (Figure 3B: lines 3-6), corresponding to the SP-6His expected in this expression system. Posttranslational modifications of an intracellular protein in *P. pastoris* cannot explain the increase of 5 kDa to generate a molecular weight of almost 25 kDa (Rath *et al*., 2009). An explanation that affects molecular weight is that the SP-6His expressed by *E. coli* is slightly lower due to the digestion of some typical proteases of the expression system. The SP-6His overexpressed by *P. pastoris* has a molecular weight of 25 kDa due to its amino acid composition that may have caused the different migration of the protein with respect to the molecular weight marker. Rath *et al*. (2009) made different observations by which a protein could show an anomalous migration in SDS-PAGE, suggesting a poor solvation of the protein with SDS to the conservation of some disulfide bridges that can maintain oligomeric forms of the protein.

On the other hand, the cloning strategy in the pQE30 vector eliminated the synthetic sequence corresponding to the ribosome binding site when the *Eco*RI/*Bam*HI endonucleases digested the pQE30 vector. Thus, the inserted sequence of the pPICZB-*SP* construct did not incorporate an alternative ribosome binding site, possibly generating a truncated peptide sequence.

Immunodetection by Western blot is considered a rapid test in this study for the preliminary identification of SP-6His tagged, but because it can be nonspecific for the histidine tag, it was purified until homogeneity for its identification by mass spectrometry (see below).

3.3 Purification of SP-6His from P. pastoris culture

Additional file 2 shows purification to homogeneity of the SP-6His, with a stronger band of 25 kDa by eluting the column with 160 mM imidazole. The soluble extract was obtained from the pellet of an expression culture (32 mL) from *P. pastoris* as described by Chávez-Cabrera *et al*. (2015) and mixed with imidazole to 20 mM. The NTA-Ni resin was contacted with the suspension of the soluble extract for 30 minutes and was loaded on the column. It was then eluted with 60-, 80-, 100-, and 160-mM imidazole. Because the protein from the samples obtained in each purification step was not quantified, it was decided to load the SDS-PAGE gel with the same amount

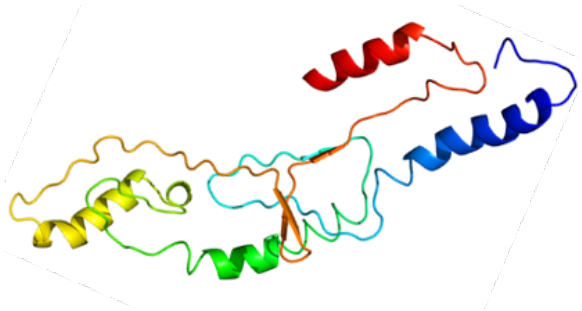

Figure 4. Prediction of the tertiary structure of SP-6His. The schematic cartoon representation of the modeling and its molecular docking of SP-6His were by Phyre2. (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index). The C-terminal of the SP-6His model (the last 41 amino acids) fits the d1ft8a2 template. The N-terminal of the SP-6His model (132 amino acids) is *ab initio*.

of protein suspension, mixing equivalent volumes of the protein suspension with the 2X denaturing buffer SDS-PAGE and boiling for 10 minutes. Finally, 20 μL of each mixture was loaded onto a 15% SDS-PAGE gel. Based on the purification train shown in Additional file 2 and the detection of the Coomassie blue stain, approximately 25 μg of synthetic protein was obtained. The poor purification performance of SP-6His suggests: first, poor cell lysis because the cells were cultured for 96 h, or the low level of the protein expression. Second, there was poor binding of SP-6His to the Ni-NTA resin because the secondary structure probably partially hides the 6His epitope even though the modeling of the SP protein indicates free ends (see figure 4). Third, the contact time between the protein and the Ni-NTA resin was insufficient. Fourth, SP-6His remains bound to Ni-NTA resin because the concentration of imidazole used during elution (160 mM imidazole) was very low (Ni-NTA Purification System). Finally, SP-6His was retained in cell membranes (cell debris) because of its hydrophobicity level (0.58), according to the ProtParam tool from Expasy (https://web.expasy.org/protparam/). Kyte and Doolittle (1982) have examined a number of membrane proteins and have identified membrane-spanning segments, as well as those hydrophobic regions that anchor certain proteins in membranes. The purification process was not intended to be optimized, as only a small amount (approx. 100 ng) is required to identify the amino acid sequence by nano-electrospray mass spectrometry.

Table 3. Peptides identified by mass spectrometry.

Peptides	Peptide sequence	Theo. MH+ [Da]
1	KFVHV	742.46102
2	FVHIVK	742.46102
3	EVIPVVVKMC	1116.61555
4	KFVPIMKNK	1104.6598
5	IHNVEKK	867.50468
6	VFVTTFPLEQKLISE	1750.96256
7	ISEEDLNSAVDHHHHHH	2013.89102
8	TVMKEVIPVVVK	1314.81742
9	EIIHNVEKK	1109.63133
10	IHNVEK	739.40971
11	NVNIVKEVKIHTVMK	1752.02004
12	VKVVPIVVFVKVTECMTK	2019.17449
13	KFVHIVKMC	1104.60565
14	VHVFFVENTINVVVK	1743.97922
15	VENHKVFIPNVNIVK	1750.00102
16	FVPIMKNKST	1164.64454
17	IHTVMKEVIP	1166.66019
18	VENTINVVVK	1114.64665
19	INVVVK	671.44504
20	IIHNVEKK	980.58874
21	IPVFVHVEVF	1166.69321
22	TINVVVKVVPIVVFVK	1753.13499
23	TFPLEQK	862.46689
24	VTECMTKIVV	1122.58973
25	FVPIMKN	848.48987
26	KIHTVMK	856.50732
27	TVVVKPEIIHNVE	1476.84206

3.4 Mass spectrometry analysis

The approximately 25 kDa band, which probably corresponds to SP-6His from *P. pastoris*, was excised from the SDS-PAGE gel for identification. Mass spectra and peptide fingerprints from trypsin digestion were analyzed by Discovery Proteome software and identified with Amanda and Sequest software search engines. Table 3 shows the identified 27 peptides, corresponding to 81% of the SP-6His coverage. It should be noted that complete trypsin digestion should generate 15 peptides. This indicates that the identified peptide sequences have one or two trypsin digestion sites inside, suggesting that the synthetic trypsinized protein was partially digested as were most of its peptide sequences. The partial digestion of SP-6His may be the reason why the theoretical coverage of 100% was not reached according to Expasy's Peptide Mass software. Thus, mass spectrometry analysis shows that the identified peptides are from SP-6His and strongly confirms the expression of a synthetic protein with a high content of essential amino acids in *P. pastoris*, due to the 81% coverage obtained. Finally, the SP-6His tertiary structure model prediction shows four alpha helix and seven beta strands (see figure 4). The 173 amino acid sequence model of SP-6His contains an ab-initio model of 132 amino acids. Since, among the protein databases used by Phyre2, the SP-6His did not match the templates of the protein models (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index).

All of these results, together with those obtained by some other working groups, indicate that the use of technological advances in DNA synthesis can achieve nucleotide sequences that encode amino acid sequences of interest with relative ease and with possible industrial and research applications, such as: in the areas of food, agroindustry, agriculture, livestock, biocatalysis, and pharmacology (Chakraborty *et al.*, 2000; Galili and Amir, 2012; Falco *et al.*, 1994; Yang *et al.*, 1989).

Conclusions

The yeast *P. pastoris*, considered a safe organism with GRAS status by the FDA, expresses a synthetic protein with a high content of essential amino acids. The synthetic protein was identified by Western blot and mass spectrometry, as highlighted below:

1. A synthetic gene encodes a protein with a high content of essential amino acids. *In-silico* analysis of synthetic protein indicates highly digestible with digestive enzymes. Also, it does not have sequences recognized as allergenic.

2. A 100% synthetic protein with 129/173 ratio of essential amino acids is expressed by *P. pastoris*.

3. The synthetic protein was purified by nickel affinity chromatography and identified by Western blot and mass spectrometry, and

4. Although it must increase current expression levels, this work represents a positive proof of concept towards producing synthetic proteins with a high content of essential amino acids. Also, this work represents a first step in creating yeast strains with a higher content of essential amino acids. These strains could be used as a single-cell protein source for livestock feed and as a food supplement for humans.

Acknowledgements

The authors thank Fernando Maldonado-García, Zoila Rosa Flores-Bustamante and Alejandro Santiago-Hernández for their expert technical assistance and for the funding of this work by the Department of

Biotechnology and Bioengineering of CINVESTAV, the Department of Microbiology of the ENCB-IPN and CONACyT for grant-aided support to CCC (Grant Number 724075) to carry out a postdoctoral research residency.

CRediT authorship contribution statement

YBA and CCC performed the experimental design, experimentation, and writing of the original draft and contributed at the same level. RM and JCCD provided the material resources and equipment and reviewed and edited the manuscript. All authors read and approved the final manuscript.

References

Angov, E., Hillier, C. J., Kincaid, R. L. and Lyon, J. A. (2008). Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PloS ONE 3(5)*, e2189. https://doi.org/10.1371/journal.pone.0002189

Baeshen, N. A., Baeshen, M. N., Sheikh, A., Bora, R. S., Ahmed, M. M., Ramadan, H. A., Saini, K. S. and Redwan, E. M. (2014). Cell factories for insulin production. *Microbial Cell Factories 13*, 141. https://doi.org/10.1186/s12934-014-0141-0

Bartolo-Aguilar, Y., Dendooven, L., Chávez-Cabrera, C., Flores-Cotera, L. B., Hidalgo-Lara, M. E., Villa-Tanaca, L. and Marsch, R. (2017). Autolysis of *Pichia pastoris* induced by cold. *AMB Express 7*, 95. https://doi.org/10.1186/s13568-017-0397-y

Bussey, H. and Umbarger, H. E. (1969). Biosynthesis of branched-chain amino acids in yeast: regulation of synthesis of the enzymes of isoleucine and valine biosynthesis. *Journal of Bacteriology 98(2)*, 623-628. https://doi.org/10.1128/jb.98.2.623-628.1969

Çelik, E. and Çalik, P. (2012). Production of recombinant proteins by yeast cells. *Biotechnology Advances 30(5)*, 1108-1118. https://doi.org/10.1016/j.biotechadv.2011.09.011

CFR - Code of Federal Regulations Title 21. (1993). U.S. Food and Drug Administration. Retrieved from https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=573.750

Chakraborty, S., Chakraborty, N. and Datta, A. (2000). Increased nutritive value of transgenic potato by expressing a nonallergenic seed albumin gene from Amaranthus hypochondriacus. *Proceedings of the National Academy of Sciences 97(7)*, 3724-3729. https://doi:10.1073/pnas.97.7.3724

Chávez-Cabrera, C., Marsch, R., Bartolo-Aguilar, Y., Flores-Bustamante, Z. R., Hidalgo-Lara, M. E., Martínez-Cárdenas, A., Cancino-Díaz, J.C., Sánchez, S. and Flores-Cotera, L. B. (2015). Molecular cloning and characterization of the ATP citrate lyase from carotenogenic yeast *Phaffia rhodozyma*. *FEMS Yeast Research 15(6)*, fov054. https://doi.org/10.1093/femsyr/fov054

Ciofalo, V., Barton, N., Kreps, J., Coats, I. and Shanahan, D. (2006). Safety evaluation of a lipase enzyme preparation, expressed in *Pichia pastoris*, intended for use in the degumming of edible vegetable oil. *Regulatory Toxicology and Pharmacology 45(1)*, 1-8. https://doi.org/10.1016/j.yrtph.2006.02.001

Cregg, J. M. (2007). *Pichia protocols*. Totowa, New Jersey: Humana Press.

Cummings, N. E., Williams, E. M., Kasza, I., Konon, E. N., Schaid, M. D., Schmidt, B. A., Poudel, C., Sherman, D. S. and Yu, D. (2017). Restoration of metabolic health by decreased consumption of branched-chain amino acids. The Journal of Physiology 596 (4), 623-645. https://doi:10.1113/JP275075

Falco, S.C., Keeler, S.J. and Rice, J.A. (1994). Synthetic storage proteins with defined structure containing programmable levels of essential amino acids for improvement of the nutritional value of plants (U.S. Patent 5,559,223). Justia Patents. https://patents.justia.com/patent/5559223

FAO. (2013). Dietary protein quality evaluation in human nutrition. Report of an FAO Expert Consultation. FAO Food and Nutrition Paper,

No. 92. Rome, FAO. Retrieved from http://www.fao.org/ag/humannutrition/35978-02317b979a686a57aa4593304ffc17f06.pdf

FAO, IFAD, UNICEF, WFP and WHO. (2018). The State of Food Security and Nutrition in the World 2018. Building climate resilience for food security and nutrition. Rome, FAO. Retrieved from http://www.fao.org/3/i9553en/i9553en.pdf

Galili, G. and Amir, R. (2012). Fortifying plants with the essential amino acids lysine and methionine to improve nutritional quality. *Plant Biotechnology Journal 11(2)*, 211-222. https://doi:10.1111/pbi.12025

Gomes, A. R., Byregowda, S. M., Veeregowda, B. M. and Balamurugan, V. (2016). An overview of heterologous expression host systems for the production of recombinant proteins. *Advances in Animal and Veterinary Sciences 4(7)*, 346-356. https://doi.org/10.14737/journal.aavs/2016/4.7.346.356

Hofacker, I. L., Priwitzer, B., Stadler, P. F. (2004). Prediction of locally stable RNA secondary structures for genome-wide surveys. *Bioinformatics 20(2)*, 186-190. https://doi.org/10.1093/bioinformatics/btg388

Institute of Medicine. (2005). Dietary reference intakes for energy, carbohydrates, fiber, fat, fatty acids, cholesterol, protein, and amino acids (macronutrients). Washington, DC: The National Academies Press. https://doi.org/10.17226/10490

Ivanciuc, O., Schein. C. H. and Braun, W. (2003). SDAP: Database and computational tools for allergenic proteins. *Nucleic Acids Research 31*, 359-362. 10.1093/nar/gkg010

Johnson, E. A. (2013). Biotechnology of non-Saccharomyces yeasts-the ascomycetes. *Applied Microbiology and Biotechnology 97*, 503-517. https://doi.org/10.1007/s00253-012-4497-y

Karbalaei, M., Rezaee, S. A. and Farsiani, H. (2020). *Pichia pastoris*: A highly successful expression system for optimal synthesis of heterologous proteins. *Journal of Cellular Physiology 235(9)*, 5867-5881. https://doi.org/10.1002/jcp.29583

Keefe, D. M. (2018). GRAS Notice GRN 737 Agency Response Letter | FDA. U.S. Food and Drug Administration. https://www.fda.gov/media/116243

Kidd, M. T. and Tillman, P. B. (2016). Key principles concerning dietary amino acid responses in broilers. *Animal Feed Science and Technology 221*, 314-322. https://doi:10.1016/j.anifeedsci.2016.05.012

Kim, H., Yoo, S. J. and Kang, H. A. (2015). Yeast synthetic biology for the production of recombinant therapeutic proteins. *FEMS Yeast Research 15*, 1-16. https://doi.org/10.1111/1567-1364.12195

Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene 361*, 13-37. https://doi.org/10.1016/j.gene.2005.06.037

Kyte, J. and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology 157(1)*, 105-132. https://doi:10.1016/0022-2836(82)90515-0

Martínez-Hernández, S., Marin-Muñoz, M., Ventura-Juarez, J. and Jauregui, J. (2019). Fed-batch cultivation and operational conditions for the production of a recombinant anti-amoebic vaccine in *Pichia pastoris* system. *Revista Mexicana de Ingeniería Química 19(2)*, 691-705. https://doi.org/10.24275/rmiq/Bio725

Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M. and Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences 101(19)*, 7287-7292. 10.1073/pnas.0401799101

Newgard, C. B., An, J., Bain, J. R., Muehlbauer, M. J., Stevens, R. D., Lien, L. F., Haqq, A. M., Shah, S. H., Arlotto, M., Slentz, C. A., Rochon, J., Gallup, D., Ilkayeva, O., Wenner, B. R., Yancy, W. S., Jr, Eisenson, H., Musante, G., Surwit, R. S., Millington, D. S., Butler, M. D., ... Svetkey, L. P. (2009). A branched-chain amino acid-related metabolic signature that differentiates obese and lean

humans and contributes to insulin resistance. *Cell Metabolism 9(4)*, 311-326. https://doi.org/10.1016/j.cmet.2009.02.002

Porro, D., Gasser, B., Fossati, T., Maurer, M., Branduardi, P., Sauer, M. and Mattanovich, D. (2011). Production of recombinant proteins and metabolites in yeasts. *Applied Microbiology and Biotechnology 89(4)*, 939-948. https://doi.org/10.1007/s00253-010-3019-z

Rashad, M. M., Moharib, S. A. and Jwanny, E. W. (1990). Yeast conversion of mango waste or methanol to single cell protein and other metabolites. *Biological Wastes 32*, 277-284. https://doi.org/10.1016/0269-7483(90)90059-2

Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., and Deber, C. M. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proceedings of the National Academy of Sciences 106(6)*, 1760-1765. https://doi:10.1073/pnas.0813167106

Rheeder, R.A.L. (2014). An exploration of synthetic biology: A preliminary Christian ethical assessment of the advantages and disadvantages of synthetic biology. *In die Skriflig 48(2)*, 1-10 pages. http://dx.doi.org/10.4102/ids.v48i2.722

Rentería-Martínez, O., Páez-Lerma, J., Rojas-Contreras, J., López-Miranda, J., Martell-Nevárez, M. and Soto-Cruz, N. (2021). Enhancing isoamyl acetate biosynthesis by *Pichia fermentans*. *Revista Mexicana de Ingeniería Química 20(2)*, 621-633. https://doi.org/10.24275/rmiq/Bio2125

Ritala, A., Häkkinen, S.T., Toivari, M. and Wiebe, M.G. (2017). Single cell protein-state-of-the-art, industrial landscape and patents 2001-2016. *Frontiers in Microbiology 8*. https://doi.org/10.3389/fmicb.2017.02009

Sambrook, J. and Russel, D. (2001). *Molecular Cloning a Laboratory Manual*. 3rd ed. Cold Spring Harbor Laboratory Press. New York.

Sánchez-Crisóstomo, M. I., Rojo-López, M. I., Sharma, A., Cancino-Diaz, J. C., Jaimes-Díaz, H., Ariza-Ortega, J. A., Madrigal-Santillán, E. and Betanzos-Cabrera, G., (2019). Construction of a synthetic protein using PCR with a high essential amino acid content for nutritional purposes. *Molecular Biology Reports 46*, 1593-1601. https://doi.org/10.1007/s11033-019-04604-1

Schulte-Herbrüggen, B., Cowlishaw, G., Homewood, K., Rowcliffe, J.M. (2017). Rural protein insufficiency in a wildlife-depleted West African farm-forest landscape. *PLoS ONE 12(12)*, e0188109. https://doi.org/10.1371/journal.pone.0188109

Thompson, C. A. (2010). FDA approves kallikrein inhibitor to treat hereditary angioedema. *American Journal of Health-System Pharmacy 67(2)*, 93. https://doi.org/10.2146/news100005

Wu, J.C., Gardner, D.P., Ozer, S., Gutell, R.R. and Ren, P. (2009). Correlation of RNA secondary structure statistics with thermodynamic stability and applications to folding. *Journal of Molecular Biology 391(4)*, 769-783. 10.1016/j.jmb.2009.06.036

Wu, F., Zhang, Q. and Wang, X. (2018). Design of adjacent transcriptional regions to tune gene expression and facilitate circuit construction. *Cell Systems 6(2)*, 206-215.e6. https://doi:10.1016/j.cels.2018.01.010

Wu, G. (2009). Amino acids: metabolism, functions, and nutrition. *Amino Acids 37*, 1-17. https://doi.org/10.1007/s00726-009-0269-0

Yang, M. S., Espinoza, N. O., Nagpala, P. G., Dodds, J. H., White, F. F., Schnorr, K. L. and Jaynes, J. M. (1989). Expression of a synthetic gene for improved protein quality in transformed potato plants. *Plant Science 64(1)*, 99-111. https://doi:10.1016/0168-9452(89)90156-8