Integrated Master's in Bioengineering

Faculdade de Engenharia da Universidade do Porto

UNIVERSITÀ
DEGLI STUDI
DI MILANO

FEUP

Cloning and expression of a new ketoreductase from the marine bacteria *Virgibacillus pantothenticus*

João Rodrigues Grilo

Dissertation for Master's Degree in Bioengineering – Specialization in Biological Engineering

Conducted at

Department of Food, Environmental and Nutritional Sciences (DeFENS), University of Milan

Supervisors: Filipe José Menezes Mergulhão (PhD)
Francesco Enzo Molinari (PhD)

September 2016

Resumo

A biocatálise emergiu como um método alternativo à síntese química de compostos intermediários destinados à indústria farmacêutica. Intensos esforços na área da investigação têm sido feitos para encontrar novas enzimas mais robustas e com capacidades inovadoras. Microrganismos extremófilos são capazes de crescer em ambientes adversos com características tais como temperaturas elevadas, elevada pressão, elevada concentração de sais. Enzimas produzidas por estes mesmos microrganismos poderão ter maior estabilidade quando submetidas a condições indústrias comparativamente às enzimas padrão utilizadas.

Um gene que codifica uma cetoredutase (kred) foi amplificado do genoma de um microrganismo marinho, *Virgibacillus pantothenticus*, que por sua vez foi isolado a partir de uma zona profunda do leste do Mediterrâneo. Este microrganismo é capaz de crescer num habitat caracterizado por uma elevada concentração de sal. O objetivo do trabalho experimental apresentado neste documento foi clonar e expressar a cetoredutase (KRED) previamente referida. Após purificação, a sua atividade seria testada em compostos bicetónicos. A enzima seria também testada sobre condições industriais tais como a presença de solventes orgânicos ou elevada concentração de substrato.

O gene kred foi clonado com sucesso com duas caudas de afinidade. A proteína foi expressa em *E. coli* como corpos de inclusão. Diferentes condições de expressão foram testadas na tentativa de obter a proteína de forma solúvel. Visto que nenhuma foi bem sucedida, um protocolo de solubilização foi seguido de modo a obter um produto recombinante solúvel. Uma solubilização parcial foi atingida, permitindo a purificação por cromatografia de afinidade (não realizada devido a limitações de tempo).

Uma análise *in silico* foi realizada para determinar o efeito das caudas de afinidade na hidrofobicidade e na exposição a solvente de KRED. Os resultados sugerem que as caudas de afinidade não provocaram um aumento da hidrofobicidade ou um decréscimo na exposição a solvente. No entanto, a presença da "His-tag" provavelmente induz alterações conformacionais na KRED tal como sugerido pela previsão de estrutura secundária.

O grupo de investigação com o qual este projeto experimental foi realizado expressou uma outra cetoredutase de forma solúvel. Uma análise *in silico* dessa proteína previu (corretamente) que esta era solúvel enquanto que a "His-tag" KRED seria altamente insolúvel. Tais resultados sugerem que outros fatores para além da hidrofobicidade e exposição a solvente são relevantes para a solubilidade de uma proteína produzida no citoplasma de *E. coli*.

Abstract

Biocatalysis has emerged as an alternative method to produce chemical intermediates for the pharmaceutical industry. Intense research efforts are being performed in this area to unveil new enzymes with more robust and innovative capabilities. Extremophile microorganisms are able to thrive in rough environments such as extreme temperatures, high pressure or high salt concentrations. Enzymes produced a by these microorganisms may have higher stability when subjected to the industrial conditions, compared to "standard" enzymes.

A ketoreductase gene (*kred*) was amplified from the genome of a marine microorganism, *Virgibacillus pantothenticus*, which was isolated from an Eastern Mediterranean basin. This microorganism is able to grow in a habitat, characterized by a very high salt concentration. The objective of this experimental work was to clone and express the mentioned ketoreductase (KRED) so that after purification its activity towards diketone compounds could be evaluated. The enzyme would further be tested under industrial conditions such as the presence of organic solvents or high substrate concentrations.

The kred was successfully cloned with two different affinity tags and the protein was overexpressed in *E. coli* as inclusion bodies. Different expression conditions were tested in order to obtain the protein in a soluble state. Since these experiments were not successful, a solubilisation procedure of the inclusion bodies was also tested. Partial solubilisation was achieved enabling protein purification by affinity chromatography (not performed due to time limitations).

An *in silico* analysis was performed to assess the effect of the affinity tails on the hydrophobicity index and solvent accessibility of the KRED protein. Results suggest that the tags did not increase the solubility or decrease the solvent accessibility. However, the presence of the His tag probably induces conformational changes in the protein as suggested by a secondary structure prediction.

In silico analysis of a related protein that was successfully expressed in the soluble state by the research group has correctly predicted that this related protein would be soluble whereas the KRED fusion protein would be highly insoluble. These results suggest that other factors besides the overall hydrophobicity or the solvent accessibility are very important to the solubility of a protein inside the *E. coli* cytoplasm.

Acknowledgments

First, I would like to thank Professor Francesco Molinari for giving me the opportunity to perform this experimental work under his supervision and together with the rest of the research group. Thank you Dr. Immacolata Serra and Danila for all the help given me throughout the project. I've learned a lot from your teachings and I will always feel in debt towards you. Finally I want to thank all the remaining members of the research group I've worked with for all the support.

To my home supervisor, Professor Filipe Mergulhão, thank you very much for all the support given me while I was living in Milan and after my return to Porto. Your enthusiastic classes pushed me to search for a project within this scientific subject.

Regarding my whole progress as a student I would like to thank to the engine that helped me surpass all the barriers, my parents. I could not have tread this academic track without the tremendous support of my girlfriend to which I will always be thankful.

As a last note I would like to thank all my friends that I've met at FEUP, to Metal&Bio and for all the good moments that we have shared.

Index

List of figures

List of tables

Notation and glossary

bp	Base pairs
CAS	Chemical Abstracts Service
DHAB	Deep Hypersaline Anoxic Basin
e.e	Enantiomeric excess
GPE	Glycidyl phenyl ether
GST	Glutathione S-transferase
His	Histidine
kred	Ketoreductase gene
KRED	Ketoreductase
LA	Linoleic acid
LNA	Linolenic acid
MaCuMBA	Marine Microorganisms: Cultivation Methods for improving their Biotechnological Applications
mRNA	Messenger ribonucleic acid
PCR	Polymerase chain reaction
PTM	Post translational modification
rRNA	Ribosomal ribonucleic acid
RPM	Rotations per minute
SOC	Super Optimal Broth
Td	Doubling time
tRNA	Transfer ribonucleic acid
Vp	*Virgibacillus pantothenticus*

Chapter 1

Introduction

1. BIOCATALYSIS

Biocatalysis is a biotechnology sector responsible for transforming different compounds through the use of biological catalysts such as isolated enzymes or whole cells. This science field increases the potential of the production of goods to meet human needs. More than daily goods, enzymes can also play a very important role in washing, environmental processes and analytical and diagnostic purposes.

There are two driving forces in the industrial and academic enzyme technology: first, the manufacture of new products or processes to meet these needs; second, the improvement of the quality of the processes for the preparation of existing products starting from different raw materials (1).

1.1 Advantages and disadvantages of biocatalysis

The employment of biocatalytic systems such as whole cells or isolated enzymes has its advantages and disadvantages (1) (2):

- Chemo-regio- and stereo-selectivity which means respectively the enzyme's capability of: interacting specifically with a functional group in the presence of other reactive groups, distinguishing between identical chemical groups located in different molecular positions and reacting with or producing molecules with a specific stereochemistry.

- Catalytic efficiency. While chemical catalysts are mainly used in concentrations of 0.1-1%, enzymatic reactions can be conducted with a biocatalyst concentration of 10^{-3}-10^{-4}%.

- Moderate reaction conditions. The majority of the enzymes can work in water around the physiological pH and mild temperatures between 20 and 40 ºC.

- Enzymes are also sustainable since they are completely biodegradable.

- Since many enzymes are stable under the same buffered conditions it is possible to set multienzymatic systems that will save reactions time and cost and can also reduce further recovery steps. A multienzymatic system can also solve the problem of having an unstable reaction intermediate.

- Enzymes can also accept a broad range of substrates that can be very different from the ones found in conventional cellular metabolism. They can also interact with a certain reaction in both directions of the equilibrium.

- However, enzymes may suffer from substrate and/or product inhibition thus requiring continuous processes.

- They are not as commercially available as chemical compounds and they are harder to scale up to an industrial process. In case they are cofactor dependent the operation cost might scale up even more as cofactors such as NAD(P)H or $FADH_2$ are very expensive. To overcome these expenses a cofactor recycling system with a secondary protein may be necessary.

- Since processes involving enzymes and whole cells are expensive and might not be profitable, recovery or reuse strategies such as immobilization must be implemented.

1.2 Enzyme thermodynamics in catalysis

Enzimes are considered catalysts since they change the Gibbs activation energy of a reaction, lowering the energy barrier between substrate and product. An enzymatic reaction is composed of two main steps, the binding step and the reaction step, as shown in Figure 1.On one hand in a non-catalyzed reaction one or more substrates are in a ground state. Once the reaction begins, the reactants surpass a transition state, that includes the point of maximum free energy, and continue to the ground state of the products. On the other hand, in an enzyme-catalyzed reaction, the free enzyme (**E**) and the free substrate (**S**) that initially were in a ground state, form a reversible enzyme-substrate complex (**ES**). The **ES** complex surpasses a lower

transition state and once activated converts the substrate into product creating the enzyme-product complex (**EP**). Finally, this complex dissociates releasing free enzyme (**E**) and free product (**P**) in their ground states (3).

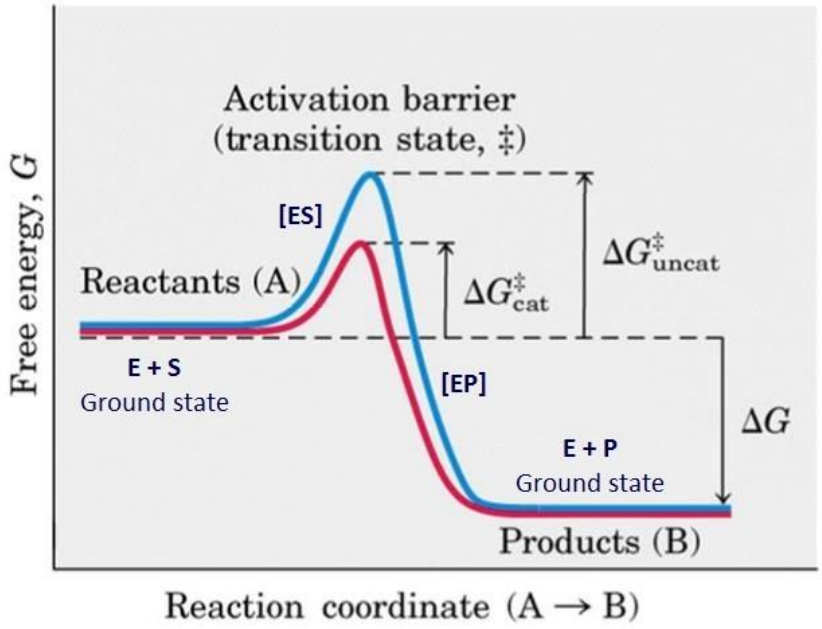


***Figure 1**. Comparison between Gibbs free energy in a catalysed VS uncatalysed reaction*

Biocatalysis can be performed using whole cells that contain the potential biocatalyst or using the biocatalyst directly in solution. Table presents the advantages and disadvantages of both procedures (2).

***Table 1**. Comparison between different biocatalysts: Advantages and disadvantages in using enzymes or whole cells (2).*

BIOCATALYST	FORM	ADVANTAGES	DISADVANTAGES
Isolated enzymes	Dissolved in water	• Simple apparatus • Simple workup • Better productivity • High substrate concentration tolerance • High catalytic activity	• Cofactor recycling necessary • Side reaction possible • Lipophilic substrates insoluble • Workup requires extraction
	Suspended in organic solvents	• Easy to perform • Simple workup • Lipophilic substrates soluble	• Reduced catalytic activity

		• Easy enzyme recovery	
Whole cells	Immobilized	• Easy enzyme recovery • Enzymatic recycling • Possibility to use continuous flow reactors	• Loss of activity during immobilization
	Growing cultures or resting cells	• No cofactor recycling necessary • High activity • Simple workup • Fewer by-products	• Expensive equipment • Tedious workup • Low productivity • Low substrate concentration tolerance • Low organic solvents tolerance • Uncontrolled side reactions
	Immobilized cells	• Possible cell re-use • Possibility to use continuous flow reactors	• Loss of activity during immobilization

In bioprocessing optimization it is necessary to know in detail the structure and other characteristics of the biocatalyst involved. Studies to obtain this type of information need to have available a considerable amount of enzyme. However, in most cases, in the original microorganism the protein of interest is produced at a basal level. Besides, in some situations, it is impossible to reproduce the microorganism that produces the protein of interest outside its natural habitat (4).

The development of recombinant DNA techniques has created methods that allow the transfer of the gene of interest into new hosts. Some of these new hosts have the ability to express and accumulate the recombinant protein, which can account for up to 30% of total cellular protein. They can also grow on cheap medium and reach a high cell density. Some of the most commonly used hosts and their valuable characteristics are listed in table 2 (5).

Nevertheless, it might be hard to foresee the outcome of a process of this type. The main drawbacks that may be encountered concern the quality of the recombinant product. In particular, the protein can undergo situations of incorrect folding. In other cases, the process fails due to the formation of inclusion bodies, which causes the aggregation of the protein making it biologically inactive. Furthermore, current legislation imposes subjecting the recombinant proteins to a series of tight controls, that intent to demonstrate the purity, the quality and the biological effectiveness of the product of interest causing a considerable increase of the production costs (6).

***Table 2**. Comparison between different expression systems (7)(8) and Protein Engineering course slides, Master's Degree in Bioengineering, FEUP.*

CHARACTERISTCS		*E. coli*	Yeast	Mammalian cells	Insect cells	Transgenic Plant cells
Cell growth		Rapid - 30 minutes	Rapid – 90 minutes	Slow – 24 hours	Slow – 18 to 24 hours	Very slow - months
Complexity of growth medium		Minimum	Minimum	Complex	Complex	Minimum
Cost of growth medium		Low	Low	High	High	Low
Cost effectiveness		Low to moderate	Low to moderate	High to very high	High	Low
Scale-up		Very good potential	Very good potential	Good potential	Good potential	Unlimited
Expression level		High	Low-high	Low-moderate	Low-high	Low
Extracellular expression		Secretion to periplasm	Secretion to medium	Secretion to medium	Secretion to medium	-
POSTTRANSLATIONAL MODIFICATIONS	**Protein folding**	Refolding usually required	Refolding may be required	Proper folding	Proper folding	Refolding may be required
	N-linked glycosylation	No	Yes	Yes	Yes	Yes
	O-linked glycosylation	No	Yes	Yes	Yes	Yes
	Phosphorylation	No	Yes	Yes	Yes	Yes
	Acetylation	No	Yes	Yes	Yes	Yes
	Acylation	No	Yes	Yes	Yes	Yes

1.3 Recombinant protein expression in E.coli

Escherichia coli (E. coli) is a Gram negative, facultative anaerobic and non-sporulating, rod-shaped bacterium that is commonly found in the lower intestine of warm- blooded organisms (endotherms) (9).

E. coli is the most used host microorganism for recombinant protein production. The main reasons that lead to this are: *E. coli* is the most well-known microorganism under different aspects; its use in the production of recombinant proteins has been refined and consolidated over time and this bacterium can be genetically manipulated with ease, allowing the development of many genetic systems. However, this microorganism has also some disadvantages: *E.coli* does not have an efficient secretion system for the release of recombinant protein into the culture medium; moreover, it also lacks the capability of establishing disulphide bonds and other post-translational modifications (PTMs), which affect protein stability and determine the efficiency of protein folding (6).

In a recombinant protein production system, the genetic manipulation of *E. coli* is a powerful way to refine and resolve problems. The genetically modified *E. coli* strains used in the present work are described in subchapter 1.8 of chapter 2 (10) (11).

2. BIOTECHNOLOGY APPLICATIONS OF MARINE MICRORGANISMS

The present technological development of food and pharmaceutical industry has been causing an intensive search for new enzymes, also known as, biocatalysts.

Marine environments, including the subsurface are believed to contain a total of approximately 3.67×10^{30} microorganisms and with approximately 71% of the earth's surface covered by the ocean, this environment represents an enormous pool of potential microbial biodiversity and exploitable biotechnology. This potential has triggered an acceleration in the study of marine microorganisms, with the aim of not only providing us with information about phylogeny, food chains and biogeochemical cycling in marine ecosystems, but also in exploiting their ability to produce novel enzymes and metabolites with potential biotechnological applications. Nevertheless, just like the terrestrial environment 99% of the marine microorganisms cannot be cultivated by conventional means (12).

Till the present date a wide variety of microorganisms that thrive in habitats characterised by extreme conditions have been identified. These microorganisms draw the attention of the scientific community since their proteins might better withstand the operational conditions of the chemical, pharmaceutical and food industry.

2.1 Marine metagenomics

Marine metagenomics is a cultivation-independent approach that can be used to find new biocatalysts from uncultured marine organisms. Metagenomics was initially used to better understand non-cultivable microorganisms such as a significant amount of marine microorganisms and their role in the global microbial ecology in different environmental niches. Nowadays, with efficient cloning vectors such as bacterial artificial chromosomes and cosmids, together with improved DNA isolation techniques and advanced screening methodologies using robotic instrumentation, the expression of large fragments of DNA and posterior screening of large clone libraries for the discovery of functional biocatalysts is possible (13).

This approach can be divided in four main steps:

I. Construction of metagenomic library;
II. Screening of metagenomic library;
III. Sequence analysis of genes;
IV. Expression of genes of interest in well-known hosts.

In the first step, genes are randomly collected by isolating total DNA. The isolated DNA is then partially digested with restriction enzymes. DNA fragments generated by the digestion are then inserted in specific circular DNA molecules called vectors. These cloned vectors are then introduced onto the expression host (e.g. *E. coli*) by heat or electric shock in a process called transformation. Each cell that that assimilated a cloned vector is called a transformant and finally the group of transformed cells constitutes the metagenomic library. The DNA inserts that were transformed may contain functional and/or phylogenetic genes. Functional genes codify proteins or biocatalysts whereas phylogenetic genes code for proteins that might not have a specific activity but are useful to establish phylogenetic relations between different organisms as for example the 16S rRNA gene.

The metagenomic library constructed needs to be screened in order to detect the cells that contain the transformed functional or phylogenetic genes, which takes us to the second step of the process. In our case of study since the functional genes we are looking for are enzymes, cell screening can be done by assaying the activity of the expression products (13). However many of the transformed genes might not expressed since their promoter regions might not be recognized by the host transcriptional machinery or be only slightly expressed due to differences in codon usage. Post-translational processes might also be needed for the enzyme to be correctly made and the host organism might not have the necessary means to establish those processes. In case the transformed genes are not expressed, techniques like *in situ* hybridization or PCR might be a good alternative for their detection.

After the screening, the third part of the metagenomic analysis steps in. The individual DNA inserts are recovered and sequenced. All sequence data is analysed through different bioinformatics tools and databases. If the properties of the biocatalysts codified by these genes are different from the one's present in these databases then we are dealing with a newly found biocatalyst that has not been reported yet. If the functional gene is located on the same DNA insert as the phylogenetic gene, then the phylogenetic gene might give information about the microbial source of the functional gene.

The final step is to choose the appropriate host and expression system for the production of the biocatalyst (13).

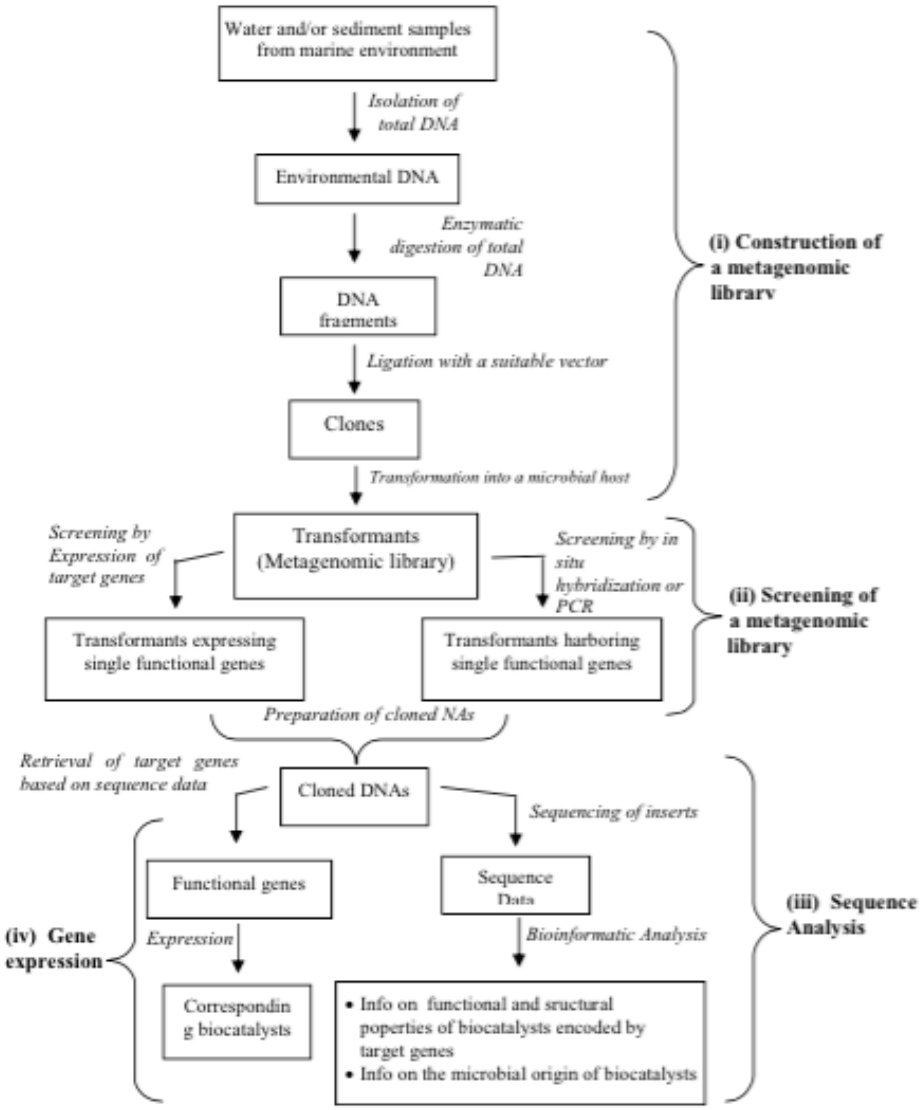


***Figure 2**. Scheme of metagenomics for the discovery of new biocatalysts from uncultured marine microorganisms, adapted from* (13).

2.2 State of the art on marine biocatalysts

Many marine microorganisms develop symbiotic relations with other organisms, those organisms are known as hosts and are commonly marine animals. Specific enzymes and metabolic pathways allow these microorganisms to adapt to other living systems (14).

Both the academic field and the industry are interested in marine extremophile microorganisms which are capable of producing specific enzymes that can withstand rough environments. These microorganisms can be classified as thermophiles or psychrophiles according to their temperature resistance and acidophiles or alkalophiles according to their pH tolerance. They can also be classified as halophiles if they thrive in high salt concentration habitats, barophiles if they are capable of withstanding high pressures and microaerophiles if they thrive in habitats with a dissolved oxygen percentage less than 21%. Microorganisms that sustain high metal concentrations or high radiation levels are classified as metalophiles and radiophiles, respectively (15). Therefore, "extremozymes" synthetized by these microorganisms are of much interest for the food and pharmaceutical industry since they have higher stability at industrial operation conditions as for example extreme pH values, high ionic strength, high pressure and presence of organic solvents.

Marine microalgae have also applications in the biotechnological field. Those organisms produce biocatalysts able to transform marine natural products. They can also serve as test organisms in ecotoxicology (16). Additionally, an increasing use of marine enzymes as biomarkers for the monitoring of the relation between marine microorganisms and environmental pollution has been observed (17). An example is the identification of proteomic signatures of exposure to marine pollutants in *Mytilus edulis*, a species of mussel (18).

Another very interesting theme to study about marine microorganisms is their capacity to catabolize marine biotoxins, especially the paralytic shellfish toxins, once commercial detoxification methods are very difficult or inexistent. Studies revealed that *Pseudoalteromonas haloplanktis* is capable of catabolizing various biotoxins (saxitoxin, neo-saxitoxin, gonyautoxins and others). The identification of the precise enzymes responsible for these catabolic processes could be very interesting in order to use them for biotransformation of bioactive natural products (19).

Nowadays, to take advantage of marine microorganism's capacity as biocatalysis it is common to use metagenomic approaches. These strategies allow us to more rapidly discover genome sequences and build clone libraries for posterior expression (12) (13) (20) (21) (22). Metagenomics provide a high amount of data about genomic diversity within the marine environment. However, it is important to identify and study the microorganisms that have specific and unique physiological characteristics. This kind of information, along with metagenomics, gives a broader understanding of the great diversity of marine microorganisms. One example of a microorganism with unique characteristics is the *Marinomonas mediterranea*,

a Gram-negative marine bacterium. These bacteria exhibit unusually high amounts of melanin and its study allowed the cloning of many genes that encoded different oxidases with great biotechnological interest (23).

2.2.1. Habitat influence on biocatalysts properties

In this present section the relation between biocatalysts properties and their producer's habitat will be presented according to scientific reports.

- ***Salt and pH tolerance***

Halophiles are organisms that need salt in their environment to live.

A very interesting point for our work is the firm relation among salt and organic solvent tolerances very often observed for halophilic enzymes because salt presence has the effect of reducing water activity (24).

In marine species, the high availability of salts might contribute to stabilizing interactions by participating in electrostatic attractions within protein complexes.

In order to know more about proteases successfully used for the production of biologically active fish protein hydrolysates, the activity of Trypsin was studied. This enzyme from the intestine of carnivorous fish smooth-hound (*Mustelus mustelus*) presented high proteolytic activity at very high (30%) concentrations of NaCl. This result demonstrates the potential of trypsin for protein hydrolysis at high salt content, and confirms its interest as a possible biotechnological tool in the fish processing and food industries (25).

Other interesting proteins for the biotechnological field are marine amylases. Saccharification, of carbohydrates is essential for preparing a process such as an alcoholic fermentation and it requires enzymes like amylases. Biocatalysts used are mainly terrestrial amylases and this has a big disadvantage: if a terrestrial enzyme is to be used, the bioprocesses of ethanol production using marine microalgae biomasses must include desalinization. Therefore, the development of a system utilizing amylase from a marine source instead of terrestrial amylases would be very beneficial. A marine bacterium that is able to produce amylase, *Pseudoalteromonas undina* NKMB 0074, was isolated and identified. This microorganism can be used in saccharification of marine microalgae producing ethanol, in saline conditions (26).

- **Hyperthermostability**

Biotechnological processes performed at elevated temperature have many advantages, such as high solubility of substrates, decreased viscosity or decreased risk of microbial contamination. Conventional proteins are completely denatured under the strict conditions in which hyperthermophilic enzymes can work. Some marine microorganisms have the capacity to

grow optimally under very high temperatures (80–108 °C). Enzymes derived from extremophilic archaea, for example, have higher stability towards heat (27).

Some works have shown that there are few mechanisms possible to impart both thermostability and interesting chemical and stereochemical features to these enzymes.

Substrate specificity and enantioselectivity are biocatalytic characteristics of an alcohol dehydrogenase from the marine hyperthermophilic archaeon *Pyrococcus furiosus* that has been carefully evaluated for major control in future biocatalytic applications. This enzyme catalyses the reduction of various ketones including alkyl and aryl ketones, and α- and ß-ketoesters. It showed also high tolerance towards organic solvents, which represents a useful feature for working with ketones possessing low solubility in aqueous buffers (28).

Thermophilic marine microorganisms that belong to *Thermotoga* sp. family possess curious glycoside hydrolases often used in transglycosylation reactions. A ß-glucosidase from the hyperthermophilic eubacteria *T. maritima* has been expressed to high levels in *E. coli*, and purified to homogeneity. This enzyme is very interesting due to its extremely thermophilicity, thermostablility and resistance to common protein denaturants as are other enzymes from the Thermotoga species. Therefore, the reported ß-glucosidase is a very good candidate as an enzyme for use in industrial applications such as alkyl glucosides production (29) (30) (31).

- ***Barophilicity***

Barophilicity is the capacity of an organism to live and thrive under high pressure conditions.

Systems under high pressure are controlled by Le Châtelier's principle: application of pressure moves an equilibrium towards the state with smaller volume, thus favouring processes for which the transition state has a smaller volume than the ground state. Definitely the major advantage of having pressure controlling as enzymatic reaction is that it is not necessary the use of special chemical agents in the reaction mixture (32).

Live organisms under the oceans must be able to resist to considerable pressure. Research on barophiles has focused, the most part of the times, on the identification of pressure-regulated operons that show the connection between pressure, high temperatures and microbial growth (33).

Recent studies (34) showed that there was no evidence of pressure adaptation in the structure or in the activity of dihydrofolate reductase from the marine barophilic bacterium *Moritella profunda*, which lives in great depth environments. Another study(35) observed that at high pressure (60 MPa) and temperature (90 °C), the number of protein bands remained unchanged when the hyperthermophilic archaeon *Thermococcus peptonophilus* was grown under high pressure and temperature but cell growth was complemented by the overproduction of specific proteins.

Biological membranes, specifically its function, and dynamic state of lipid components are closely related. There is a striking correlation between growth at high pressure and fatty acid unsaturation index. An increase in unsaturated fatty acids leads to highly disordered phospholipid bilayers, which makes the membrane feasible against pressure effects. In this context, barophilic bacterial lipid biochemistry is a very interesting aspect and enzymes acting in metabolic routes under these extreme conditions could be very interesting for biocatalytic applications (36).

- ***Cold adaptivity***

Enzymes that act at low temperatures (Psychrophilic enzymes) have a great potential for the biotechnological field. These biocatalysts could be used as additives in the industries of foods and detergents, or in bioremediation processes: they could reduce the risk of microbial contamination, minimize energy consumption or avoid the high temperature instability of reactants or products. Psychrophilic enzymes have some advantages: their high specific activity (allowing reduction of the amount of enzyme needed) and their easy inactivation, which prevents prolonged action of the enzymes when undesired.

In order to use cold-adapted microorganisms as a source of exploitable enzymes it is very important the isolation and characterization of bacteria that are able to remove lipids at low temperatures (37).

A recent study (38) was based on a cold-adapted esterase of a novel marine isolate, *Pseudoalteromonas arctica*. In particular, gene cloning, enzyme purification and characterization were reported. This esterase displayed broad substrate specificity for short-chain fatty acid esters (C2–C8) and showed an optimum pH of 7.5 and optimum temperature of 25 °C, but with more than 50% retained activity at the freezing point of water. Additionally, the enzyme was capable of hydrolysing esters of medical relevance such as non-steroidal anti-inflammatory drugs (naproxen, ketoprofen, and ibuprofen).

- ***Habitat-related chemical and stereochemical properties***

Stereoselective reactions are recognized for producing an unbalanced mixture of stereoisomers. They can either be enantioselective when the two products are enantiomers or diastereoselective when the two products are diastereoisomers. In this sub-chapter some enantioselective reactions will be presented. The enantiomeric excess measures the composition of the enantiomeric mixture and is calculated using the following formula

$$e.e(\%) = \frac{R-S}{R+S} \times 100 \qquad [1]$$

$$R = Concentration\ of\ the\ (R)\ enantiomer$$

$$S = Concentration\ of\ the\ (S) enantiomer$$

An extremely enantioselective epoxide hydrolase (REH) was found in the marine bacterium *Rhodobacterales bacterium* HTCC2654. The enantioselective hydrolysis of a 29.2 mM racemic mixture of glycidyl phenyl ether (GPE) using 20 µg of purified REH was carried out. The (S)-GPE hydrolysis rate was significantly faster than (R)-GPE one, resulting in a calculated e.e. of 99.9% of (R)-GPE within 22 minutes of reaction. The high optical purity of (R)-GPE was confirmed by the measurement of the reaction products concentration since it revealed a much higher concentration of (S)-3-phenoxy-1, 2-propanediol. The authors affirmed that according to their knowledge this was the highest enantioselectivity value amongst the known epoxide hydrolases for the enantioselective hydrolysis of GPE (39).

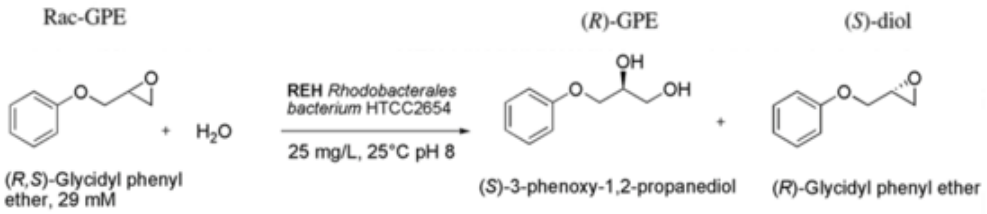


***Figure 3**. The enantioselective hydrolysis of the purified REH from Rhodobacteral bacterium HTCC2654 toward racemic GPE. Adapted from* **(39)**.

Another study found epoxide hydrolase activity in an isolated marine microorganism, *Sphingomonas echinoides* EH-983. A racemic mixture of 40 mM of styrene oxide was supplied and hydrolysed by whole cells of *Sphingomonas echinoides* EH-983. Enantiopure (e.e. 99%) (S)-styrene oxide was produced with a yield of 21.3% within 180 minutes of reaction. The EH-983 had a significant affinity with the (R)-styrene oxide forming its corresponding R-diol at a very fast pace (40).

Polycyclic aromatic hydrocarbons (PAHs) are a class of organic pollutants that exist in the marine environment mainly due to human activities. Basically, these pollutants are hydrocarbons with two or more chained benzene rings. These chemicals have been an environmental concern due to their persistence, bioaccumulation and carcinogenicity (41). *E. coli* was used to express an aromatic dihydroxylating dioxygenase gene from marine bacteria *Nocardioides* sp. KP7. It was found by the authors that this dioxygenase had a substrate preference for various mono- or di-substituted naphthalenes such as dimethylnaphthalenes, specifically the 1,4-dimethylnapthalene that was frequently converted to 1,4-dihydroxymethylnaphtalene. These hydroxylated products are industrially useful as starting materials in the chemical industry (42)(43).

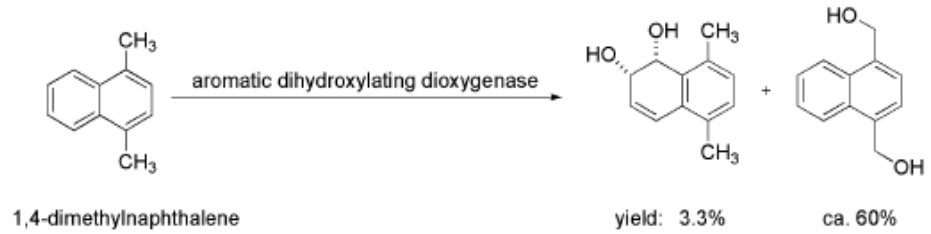


***Figure 4**. Reaction of transformation of 1,4-dimethylnaphthalene using the cells of Escherichia coli expressing aromatic dihydroxylating dioxygenase* (43).

Alcanivorax borkumensis represents an alternative to current chemical oxygenating processes of normal alkanes (n-alkanes). This microorganism has a solid position on the marine carbon cycle and is involved in biodegradation of oil spills. Miri *et al.* discovered two membrane alkane hydroxylases (AlkB1 and AlkB2) capable of hydroxylating both linear and branched alkanes. We stand before two biocatalysts that may arouse great interest in the chemical industry for the biotransformation of fatty acids and alkanes (44).

Still in the scene of carbohydrate biotransformation a newly isolated marine bacterium, *Agarivorans* sp., was found to produce an endo-type ß-agarase. The enzyme was expressed extracellularly in *E. coli* and was capable of hydrolysing the ß-1,4-glycosidic linkages of agarose yielding neoagarotetraose and neoagarohexaose as the main products. It also retained more than 95% of its activity when it was incubated within 3 pH values away from its optimum pH. The characteristics mentioned previously make this biocatalyst of potential use, especially in the cosmetics industry (45).

Besides di-*myo*-inositol-1,3'-phosphate (DIP) which is a widespread solute over marine hyperthermophile microorganisms, two newly ß-1,2-mannosides were identified on *Thermotoga maritima*, 2-(*O*-ß-D-mannosyl)-di-myo-inositol- 1,3-phosphate (MDIP) and 2-(*O*-ß-D-mannosyl-1,2-*O*-ß-D-mannosyl)-di-myo-inositol-1,3-phosphate (MMDIP). *T. maritima* cells grown *in vitro* under heat stress revealed a high accumulation of MDIP that accounted for 43% of the total pool of solutes per cell. The synthesis of MDIP involved the transfer of the mannosyl group from GDP-mannose to DIP in a single-step reaction catalyzed by MDIP synthase. Putative genes for MDIP synthase were identified in the genome of *T. maritima*, and were recombinantly expressed in *E. coli*. The enzyme showed maximum activity at 95 ºC with a 16 mM K_m value for DIP. Therefore, a new glycosyltransferase is now available that can sustain biotransformation processes at high temperatures (46).

A newly isolated strain from a marine environment, identified as *Yarrowia lipolytica*CL180 revealed esterase (*yl180*) activity on anofloxacin ester substrate. The authors of this study verified that the *yl180*hydrolized preferentially the *S*-enantiomer of a racemic mixture of ofloxacin ester resulting in the formation of levofloxacin (*S*-isomer of ofloxacin). Levofloxacin, S-isomer of ofloxacin, shows a broad spectrum of antibacterial activity against both gram-positive and gram-negative bacteria and is typically two times more potent than racemic

ofloxacin. Nevertheless, the measured enantiomeric excess was 52.1% and this value might not be good enough to fit the actual market but further improvements will be done. Additionally this enzyme showed a 40% maximal activity at 10 ºC which is a thrilling feature for heat-labile chemical production (47).

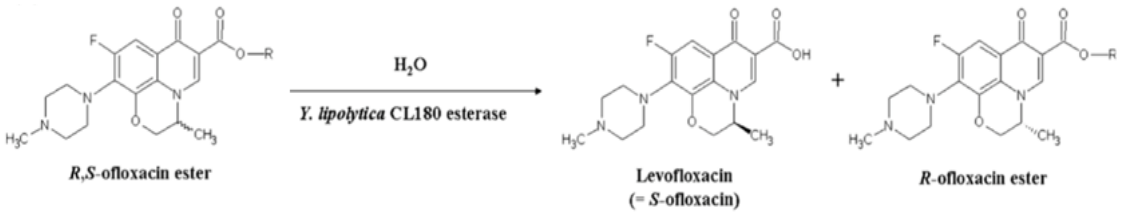


***Figure 5**. Reaction scheme of the enantioselective resolution of Y. lipolytica CL180 esterase toward R, S-ofloxacin ester* (47).

Akakabe*et al.*, incubated a crude enzyme, more specifically a lypoxygenase, from the marine algae *Ulva conglobata* with linoleic (LA) and linolenic acid (LNA) as substrates. This enzymatic study was assessed since the authors had previously done a GC-MS on a prepared essential oil from the algae and found oxidized forms of the fatty acids previously mentioned. The corresponding (R)-9-hydroperoxy-(10E, 12Z)-10, 12-octadecadienoic acid [(R)-9-HPODE] and (R)-9-hydroperoxy-(10E, 12Z, 15Z)-10, 12, 15-octadecatrienoic acid [(R)-9-HPOTrE] were formed with a high enantiomeric excess (>99%), respectively. A significant number of studies of this type of enzyme have already been made on terrestrial plants enhancing for example their bleaching capability (48). However, the regio- and stereo- selective biocatalytic action of the enzyme reported by Akakabe *et al.* differs from those found in plants. This is the first time that (R)-9-and (R)-9-HPOTrE were found in marine algae (49).

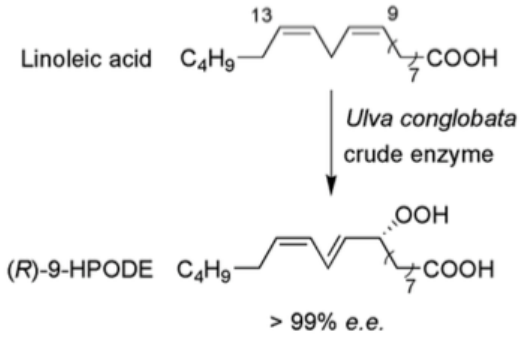


***Figure 6**. Ulva conglobata crude enzyme reaction forming (R)-9-hydroperoxy-(10E,12Z)-10,12-octadecadienoic acid ((R)-9-HPODE) (43).*

Still in the non-microorganism scene, Andreotti *et al.* synthetized ß-galactosyl derivatives of antiviral and anticancer nucleosides thanks to a purified ß-galactosidase from the hepatopancreas of *Aplysia fasciata*, a sea hare (50). The enzyme regioselectivity was extremely

high, since in all reactions only the product of galactosylation in 5′ position of the nucleoside was observed. 5′-*O*- ß-galactosyl-5-fluorouridine, the galactosylated derivative of the anticancer drug fluorouridine, was synthesized with a 60% yield, and 5′-*O*- ß-galactosyl-3′ -azido-3′-deoxythymidine, the derivative of the anti-HIV drug, was obtained in 43% yield. This was the first report about a glycosyl hydrolase used for the modification of nucleosides with convenient yields when compared to enzymes from other sources. The reaction products might be very useful for the pharmaceutical industry regarding the treatment of harmful non-curable viral diseases and cancer (50).

2.3 New ketoreductase from the marine bacteria Virgibacillus pantothenticus

Within the scientific environment presented, between February and July of 2016, a research work was developed on the cloning and expression of a gene from the marine bacteria *Virgibacillus pantothenticus (Vp)*. Under the patronage of an European project named MaCuMBA: Marine Microorganisms: Cultivation Methods for Improving their Biotechnological Applications, different isolated strains from a water-brine interface of a Deep Hypersaline Anoxic Basin (DHAB) were screened. Amongst the isolated microorganisms, the genome of the *Virgibacillus pantothenticus* strain was sequenced by a next generation sequencing method. From its genome, a gene codifying a ketoreductase, **kred**, was retrieved. The enzyme (**KRED**) belongs to the family of ketoreductases (also known as alcohol dehydrogenases) with the EC number 1.1.1. Ketoreductases catalyse reactions like the reduction of ketone functional groups or the oxidation of alcohols (42).

Different ketoreductases are currently used in the chemical and pharmaceutical industry. Nevertheless, the motif that drove this research was the possibility that this KRED could operate with a better performance under industrial conditions such as the presence of organic solvents or high ionic strength solutions that might be similar to the ones verified on the DHAB's.

Deep hypersaline anoxic basins (DHABs) in the Eastern Mediterranean Sea, located at depths of more than 3000 m below sea level, result from the dissolution of outcropping ancient subterranean salt deposits from the Messinian salinity crisis (late Miocene period, > 5 million years ago). Posteriorly, in the Zanclean age, water from the Atlantic rushed in through the actual Strait of Gibraltar covering again the exposed dissected areas. Due to the high densities of the brines (up to 1.23 $kg.m^{-3}$), mixing of these water masses with overlying deep-sea water (average density: 1.03 $kg.m^{-3}$) is restricted, resulting in anoxic conditions in these brines. An interface (halocline: depending on the basin, typically 1 to 3 m thick) separates the anoxic brine from the normoxic and normal saline deep-sea water (51) (52).

3. GOAL AND MOTIVATION

The **main goal** of this experimental work was to **clone** and **express** a **new ketoreductase** from ***Virgibacillus pantothenticus*** for further test its activity. Once KRED was obtained, its activity would be tested against simple diketone compounds such as benzyl or dibenzoylmethane. The enzyme's activity would also be tested under industrial conditions such as working with high substrate concentration or in the presence of high ionic strength solutions as well as organic solvents to evaluate its potential as a biocatalyst.

Chapter 2

Materials and Methods

1. CLONING

In this section, a first cloning approach is described in order to clone the kred gene in frame with a 6x his-tag sequence. After being unsuccessful and regarding the available time for the project, a cloning kit from Invitrogen™ (Champion™ pET Directional TOPO® Expression Kits) was used to speed up the project. The kred gene was also cloned in a pGEX-4T-1 vector that coded a different affinity tag (glutathione S-transferase, GST).

1.1 Expression Vectors

Throughout the project, three different vectors were used to clone the kred gene. First an attempt to clone kred in pET-14b vector was made. The vector used in the cloning kit was the pET100 TOPO vector which codified a histidine affinity tag to be fused with the protein. To evaluate the expression of the KRED with a different affinity tag, the cloning was also done on the pGEX-4T-1 vector. The gene was cloned in frame with a Glutathione S-transferase (GST) tag.

- **pET-14b:** This vector provides resistance to ampicillin (ampR) and contains a high-copy number pBR322 origin of replication. The multiple cloning site works under the command of a T7 promoter (strong) region. The restriction sites chosen for the cloning were XhoI and BamHI.pET-14b also presents a sequence coding a 6xHis affinity tag prior and in frame with the multiple cloning site. In between these two regions a thrombin recognition and cleavage site is codified for later excision of the purification tag. This vector's map is presented in Figure 7.

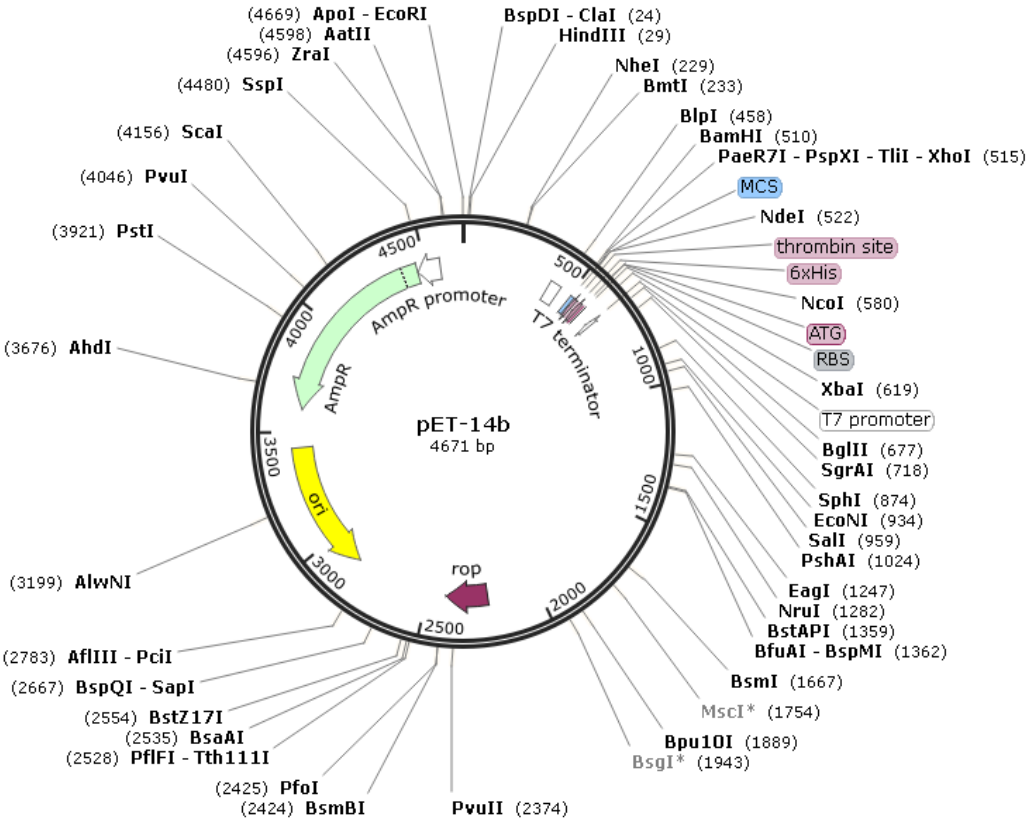


***Figure 7**. pET-14b vector. Picture obtained using SnapGene® software*

- **pET-100 TOPO:** This vector provides resistance to ampicillin (ampR) and contains a high-copy number pBR322 origin of replication. The multiple cloning site works under the command of a T7 promoter region. Transcription of any cloned gene is controlled by the lac operator. pET-100 TOPO also presents a sequence coding a 6xHis affinity tag prior and in frame with the multiple cloning site. In between these two regions an enterokinase recognition and cleavage site is codified for later excision of the purification tag. This vector's map is presented in Figure 8.

- **pGEX-4T-1:** This vector provides resistance to ampicillin (ampR) and contains a high-copy number pBR322 origin of replication. The multiple cloning site works under the command of a tac promoter region. The restriction sites chosen for the cloning were BamHI and EcoRI. Transcription of any cloned gene is controlled by the lac operator. pGEX-4T-1 also presents a sequence coding a glutathione S-transferase (GST) affinity tag prior and in frame with the multiple cloning site. In between these two regions a

thrombin recognition and cleavage site is codified for later excision of the purification tag. This vector's map is presented in Figure 9.

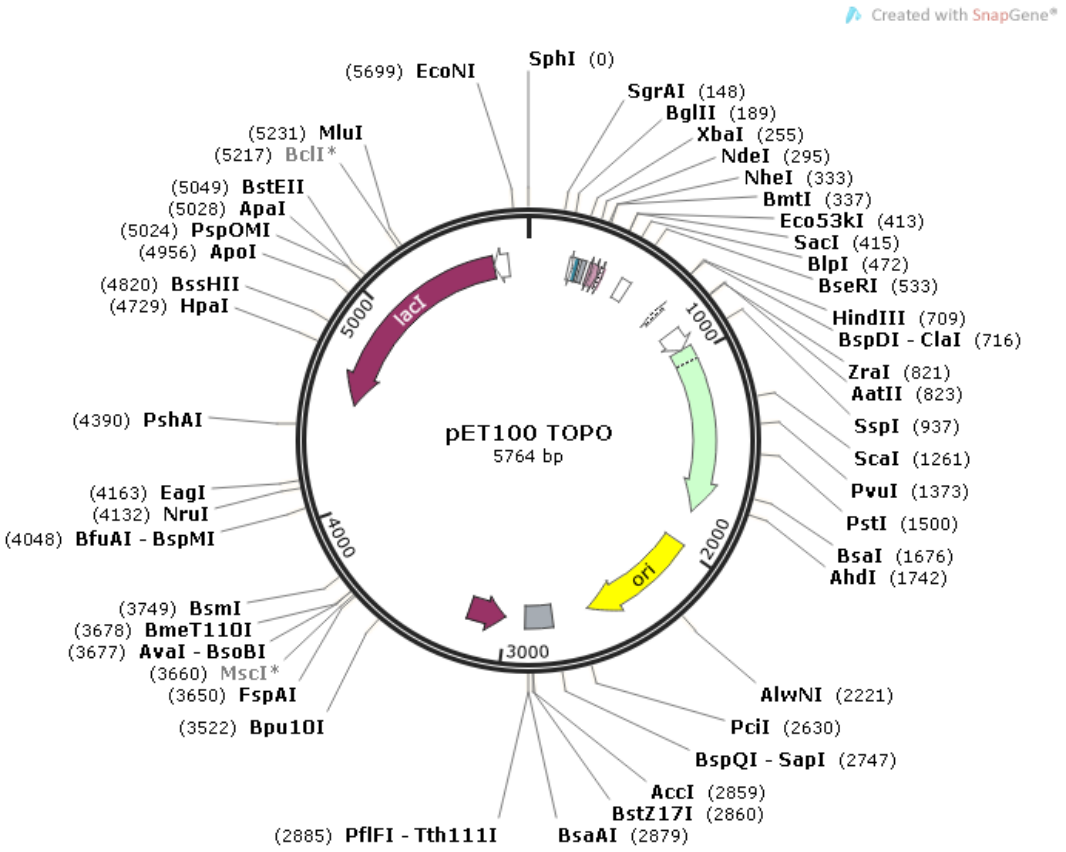


***Figure 9**.pET100 vector. Picture obtained using SnapGene® software*

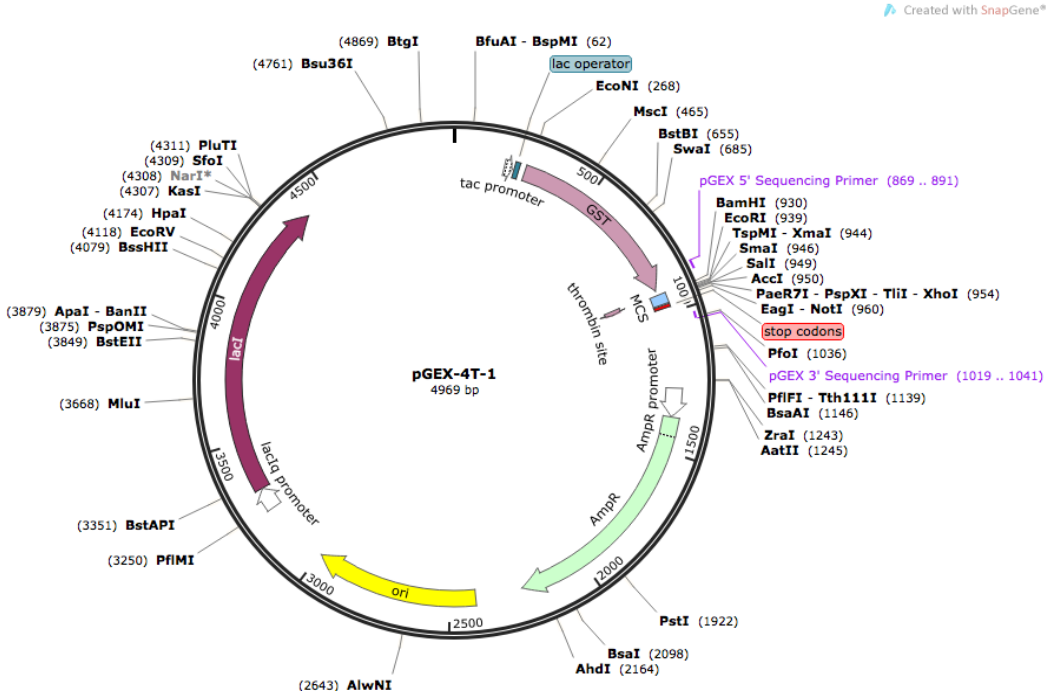


***Figure 8**. Vector pGEX-4T-1. Picture obtained using SnapGene® software*

1.2 Amplification of the kred from the genome of Vp

Note: Please take into account that all enzymatic digestions, ligation reactions and PCRs were done in an Eppendorf flexlid Mastercycler® nexus gradient.

The gene of interest (kred) was amplified through PCR with two different sets of primers, whose sequences are described in Table 3. The first set of primers codified restriction sites for XhoI and BamHI directed to a cloning on the pET-14b vector. The second set of primers codified restriction sites for BamHI and EcoRI directed to a cloning on the pGEX-4T-1 vector.

***Table 3**. Primers used in the kred amplification from the Vp genome (restriction sites indicated in capital letters)*

Primers	Sequence
XhoI_pET-14b _FW	caccCTCGAGatggattatgcaattgttacag
BamHI_pET-14b_RV	agcaGGATCCttaagataaataatcacgtgcataata
BamHI_pGEX-4T-1_FW	caccGGATCCatggattatgcaattgttacag
EcoRI_pGEX-4T-1_RV	agtcGAATTCttaagataaataatcacgtgcataa

Both 5' primers (XhoI_pET-14b and BamHI_pGEX-4T-1) have a 5'-cacc as a preventive measure that allowed the use of a cloning kit that is described further down in this document.

The DNA polymerase used was the Phusion High-Fidelity (Thermo Fisher Scientific Inc.) to minimize errors during the polymerization and guarantee a PCR product without mutations.

The reaction mix was done according to Table 4, and the PCR conditions are explained in Table 5.

***Table 4**. Reaction mix for kred amplification PCR*

Reagent	Volume (µL)
Milli-Q H2O	11,7
High-Fidelity Buffer (5x)	4
dNTP's (2,5 mM)	1,6
FW primer (10 µM)	1
RV primer (10 µM)	1
Vp genomic DNA (100 ng/µL)	0,5
Phusion High-Fidelity DNA polymerase (2U/µL)	0,2

***Table 5**. Conditions of kred amplification PCR*

STEP	TEMPERATURE	TIME	CYCLES
1 (Initiation/Denaturation)	98 °C	7 minutes	1
2(Denaturation)	98 °C	10 seconds	27
3(Annealing)	Tm gradient °C {59, 60, 61, 62, 63}	20 seconds	
4(Extension)	72 °C	25 seconds	
5(Final elongation)	72 °C	7 minutes	1

1.3 PCR clean-up kit (according to user's manual)

The kit used for this clean-up step was ISOLATE II PCR and Gel Kit, from Bioline. This kit is specialized to isolate extremely pure DNA by removing contaminants such as primers, enzymes, mineral oil, PCR additives and detergents.

The protocol used for PCR clean-up is described below:

1. **Sample preparation**
   For volumes <30 µL, adjust volume to 50–100 µL with water.
   Mix 1 volume of sample with 2 volumes of Binding Buffer CB.
2. **Bind DNA**
   Place an ISOLATE II PCR and Gel Column in a 2 mL Collection Tube and load sample.
   Centrifuge 30s at 11 000 x g and discard flow-through.
   Reuse collection tube for step 3.
3. **Wash silica membrane**
   Add 700 µL Wash Buffer CW to ISOLATE II PCR and Gel Column.
   Centrifuge 30s at 11 000 x g.
   Discard flow-through and place column back into Collection Tube.
   Recommended: Repeat washing step to minimize chaotropic salt carry-over.
4. **Dry silica membrane**
   Centrifuge 1 min at 11 000 x g, to remove residual ethanol.
   Place ISOLATE II PCR and Gel Column in a 1.5 mL microcentrifuge tube (not supplied).
5. **Elute DNA**
   Add 15-30 µL Elution Buffer C directly onto silica membrane.
   Incubate at room temperature for 1 min.
   Centrifuge 1 min at 11 000 x g.

For more information about the kit please visit the following link: http://www.bioline.com/sg/downloads/dl/file/id/878/isolate_ii_pcr_and_gel_kit_product_manual.pdf .

1.4 Enzymatic digestions

1.4.1 Enzymatic digestion of pET-14b

Different digesting conditions were tried throughout the scientific work. The standard conditions used for digesting approximately 500 ng of pET-14b with an average concentration of 87.220 ng/µL are described in table 6.

***Table 6**. Enzymatic digestion of pET-14b reaction mix*

Reagent	Volume (µL)
Milli-Q H2O	10
FD Buffer (10x) or FD Green Buffer (10x)	2
DNA	6
FD XhoI	1
FD BamHI	1

The digestion mix was incubated at 37 ºC for 1 hour followed by an enzyme inactivation at 80 ºC for 20 minutes.

1.4.2 Enzimatic digestion of pGEX-4T-1

Different digesting conditions were tried throughout the scientific work. The standard conditions used for digesting approximately 500 ng of pGEX-4T-1 with an average concentration of 68.0 ng/µL are described in Table 7.

***Table 7**. Enzymatic digestion of pGEX-4T-1 reaction mix*

Reagent	Volume (µL)
Milli-Q H2O	8.6
FD Buffer (10x) or FD Green Buffer (10x)	2
DNA	7.4
FD XhoI	1
FD BamHI	1

The digestion mix was incubated at 37 ºC for 1 hour followed by an enzyme inactivation at 80 ºC for 20 minutes.

1.4.3 Enzimatic digestion of amplified kred gene

After using the PCR clean-up kit on the amplified kred gene, the gene was digested with the matching restriction enzymes according to the primers used on the gene PCR. The standard conditions used for digesting the amplified gene are described in Table 8.

***Table 8**. Conditions used for digesting the kred amplified gene*

Reagent	Volume (µL)	Volume (µL)
-	**kred for pET-14b undigested(93.1 ng/µL)**	**kred for pGEX-4T-1 undigested (72.9 ng/µL)**
Milli-Q H2O	14.93	14.61
FD Buffer (10x)	2	2
DNA (100 ng)	1.07	1.39
FD XhoI	1	-
FD BamHI	1	1
FD EcoRI	-	1

The digestion mix was incubated at 37 ºC for 1 hour followed by an enzyme inactivation at 80 ºC for 20 minutes.

1.5 Dephosphorylation

To prevent self-ligation of the vector due to compatible ends of the digested sites, the digested vectors were dephosphorylated before any ligation step. By removing both 5' phosphate groups of the digested vector, it will not be able to self-ligate. This is achieved by using an alkaline phosphatase that will be able to break the ester bond between the phosphate group and the adjacent 5' nucleoside.

Since the alkaline phosphatase used is provided by ThermoFisher Scientific Inc. and is compatible with the Fast Digest Buffer (10x), all dephosphorylating reactions were done together with the digestion of the vectors. Therefore, this step was simplified by adding 1 µL of FastAP Thermosensitive Alkaline Phosphatase (1 U/µL)to the digestion mix and filling up the mix volume with Milli-Q H_2O till it reached a total volume of 20 µL.

1.6 Ligation

To ligate the digested gene to the digested and dephosphorylated vector the T4 DNA Ligase was utilized.

Different mixes with different gene/vector ratios were produced by floating the ng of gene used against a constant amount of vector (25 ng) (Table 9).To obtain the 1:1 ratio, the following calculus was done, taking in consideration the base pair number of the gene and the vector.

$$\frac{500\ ng}{20\ \mu l} = 25\ ng/\mu L\ vector \qquad [2]$$

$$\frac{100\ ng}{20\ \mu l} = 5\ ng/\mu L\ gene \qquad [3]$$

$$\frac{759\ bp}{4969\ bp} \times 25\ ng\ of\ vector = 3.82\ ng\ of\ gene \qquad [4]$$

$$\frac{3{,}82\ ng}{5\ ng/\mu l} = 0.77\ \mu L\ of\ insert \qquad [5]$$

***Table 9**. Ligation reaction mixes*

Reagent	Volume (µL)				
	Ratios				
	1:1	**1:3**	**1:5**	**1:10**	**Control**
Milli-Q H2O	15.73	14.19	12.65	8.8	16.5
T4 Buffer (10x)	2	2	2	2	2
Vector (25 ng/µL)	1	1	1	1	1
Insert (5 ng/µL)	0.77	2.31	3.85	7.7	-
T4 DNA Ligase (5U/µL)	0.5	0.5	0.5	0.5	0.5

The ligation mixes were incubated at 22 ºC for 1.5 hours followed by an enzyme inactivation at 65 ºC during 10 minutes.

1.7 Transformation

All ligation products were transformed into the *E. coli* DH5α™ transformation strain.

The protocol used for the transformation step is described below:

- Add the DNA (1 µL of plasmid construct or 10 µL of ligated DNA mix) to an aliquot containing 200 µL of competent cells. **Do not mix by pipetting since competent cells are very sensitive to shear stress**. Mix gently using the tip of your finger.
- Incubate in ice for 30 min.
- Incubate at 42 ºC for 1 min.
- Add 800 µL of fresh sterilized LB medium.
- Incubate at 37 ºC 150 RPM for one hour.
- Centrifuge at 2300 x g for 3 min and discard 900 µL of supernatant.
- Resuspend the pellet.
- Split the remaining in two (10 µL and 90 µL) LB agar plates containing the appropriate selective marker.
- Incubate the plates overnight at 37 ºC.

1.8 Strains

During the experimental project, a total of three different *E. coli* strains were utilized. A brief description of each strain was done in this sub-chapter.

- **DH5α™**

Chemically competent DH5α™ *E. coli* is the most commonly used strain for cloning procedures. Its genotype is known for enhancing plasmid preservation.

Genotype: F- φ80lacZΔM15 Δ(lacZYA-argF) U169 deoR recA1 endA1 hsdR17 (rk-, mk+) phoA supE44 λ- thi-1 gyrA96 relA1.

Some of its highlight features are: The endA1 mutation disables the intracellular endonuclease I which may degrade plasmid DNA therefore increasing the yield of plasmid extraction procedures. The recA1 mutation inactivates a RecA protein that induce homologous recombination between the insert and other single stranded DNA. This mutation guarantees a more stable insert without substitutions or insertions. The lac Z gene is partially mutated and is known for lacZΔM15. The strain produces an inactive form of the ß-galactosidase and is therefore suitable for blue/white screening with X-gal.

After being transformed with the recombinant plasmids utilized in this project, the strain was always inoculated with 100 µg/mL of ampicillin.

- **TOP10**

One Shot® TOP10 chemically competent *E. coli* are similar to the DH5α™ and have the same described highlighted features. This strain is part of strains the product Champion™ pET Directional TOPO® Expression Kits.

Genotype: F- mcrAΔ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu)7697 galUgalKrpsL (StrR) endA1 nupG.

After being transformed with the recombinant plasmids utilized in this project, the strain was always inoculated with 100 µg/mL of ampicillin.

- **BL21(DE3) STAR™**

The E. coli BL21 (DE3) STAR™ is an expression strain optimized for the expression of non-toxic recombinant proteins. This strain contains a specifically designed genome that enhances mRNA stability and protein yield. It is also optimized for the use with low copy number of T7 promoter based plasmids since it contains the λDE3 lysogen which has a gene controlled by the lacUV5 promoter that encodes a T7 RNA polymerase. mRNA stability and protein yield is also one of the highlights of this strain. It contains a mutation in the RNaseE gene (rne131). The mutated gene encodes an RNaseE that lacks 477 C-Terminal amino acids which are normally responsible for the degradation of mRNA, therefore increasing the stability of the transcripted mRNAs and increasing the protein yield. Two common types of proteases, from the Lon and OmpT families, that can be found in bacteria, are also absent on the BL21 (DE3) STAR.

Genotype: F- *ompT* *hsd*S_B (r_B-, m_B-) galdcmrne131 (DE3).

After being transformed with the recombinant plasmids utilized in this project, the strain was always inoculated with 100 µg/mL of ampicillin.

- **BL21-CodonPlus™(DE3)-RIPL (CODON+)**

Just like the BL21 STAR (DE3)™, the *E. coli* BL21-CodonPlus™(DE3)-RIPL contains the λDE3 lysogen which is convenient for the T7 promoter recombinant expression. Withal it features a ColE1-compatible, pACYC-based plasmid containing extra copies of the argU(AGA, AGG) and proL(CCC) tRNA genes and a ColE1- and pACYC-compatible pSC101- based plasmid containing extra copies of the argU(AGA, AGG), ileY(AUA), and leuW(CUA) tRNA genes. Availability of such plasmids solves the problem of recombinant expression in organisms that have a high AT or GC content in their genomes. These plasmids also assign resistance to streptomycin and chloramphenicol.

After being transformed with the recombinant plasmids utilized in this project, the strain was always inoculated with 100 µg/mL of ampicillin, 75 µg/mL of streptomycin and 50 µg/mL of chloramphenicol.

Genotype: *E. coli* B F- ompThsdS(rB- mB -) dcm+Tetr gal λ(DE3) endAHte [argUproLCamr] [argUileYleuW Strep/Specr].

1.9 Method to prepare competent cells

In order to have enough aliquots available with chemically competent transformation and expression cells the following procedure was performed at the beginning of the project: Inoculate the desired strain overnight in 10 mL of LB at 37ºC 150 RPM

1. Take the pre-inoculum out of the incubator and measure its $OD_{600\ nm}$.
2. Perform the required calculus to inoculate the necessary volume to obtain a 0.1 $OD_{600\ nm}$.
3. Add the calculated volume of pre-inoculum to the 80 mL LB flask and incubate it at 37 ºC 150 RPM.
4. Measure the $OD_{600\ nm}$ often considering a Td of 20 min and retrieve the flask once an $OD_{600\ nm}$ 0.4 - 0.5 is reached.
5. Split the culture onto two 50 mL falcon tubes.
6. Centrifuge the tubes for 10 min at 5000 x g and 4ºC.
7. Discard the supernatant and resuspend the pellet in 2 mL of solution A.
8. Re-fill the tubes with solution A until a 20 mL volume is reached.
9. Centrifuge the tubes for 10 min at 5000 x g and 4ºC.
10. Discard the supernatant and resuspend the pellet in 2 mL of solution B.
11. Re-fill the tubes with solution B until a 20 mL volume is reached.
12. Incubate the tubes in ice for 30 min.
13. Centrifuge the tubes for 10 min at 5000 x g and 4ºC.
14. Discard the supernatant and resuspend the pellet in 4 mL of solution B with 10% of glycerol (7mL of solution B + 1mL of glycerol 87%).
15. Split the cell stock in 200 µL aliquots and preserve them at -80 ºC.

The recipes used to prepare solutions A and B are described in Table 10.

***Table 10**. Recipes to solutions A and B, used in the preparation of competent cells*

Solution A	Solution B
10 mM MOPS pH 6.5	10 mM MOPS pH 6.5
10 mM RbCl	50 mM $CaCl_2$
-	10 mM RbCl

1.10 Champion™ pET Directional TOPO® Expression Kits

As mentioned previously, a second attempt was made in order to clone the kred gene in frame with a 6x His tag sequence. In this attempt the Champion™ pET-100 Directional TOPO® Expression Kit was used. This cloning kit is described in the following section.

1.10.1 TOPO® Cloning Reaction and Transformation (according to user's manual)

Topoisomerase I from *Vaccinia virus* is an enzyme that acts both as an endonuclease and as a ligase. It binds to double-stranded DNA but it acts on single-stranded DNA (ligase). It cleaves the phosphodiester backbone after the 5'-CCCTT and unwinds the strand one turn. The energy released from the cleavage is conserved by the formation of a covalent bond between the 3' phosphate of the cleaved strand and a tyrosyl residue (Tyr-274) of the topoisomerase I. This cleavage produces an overhang GTGG that invades the 5′-CACC end of the PCR product, anneals to the added bases, and stabilizes the PCR product in the correct orientation.

Preparation for the TOPO cloning was done by calculating the gene/vector ratio. Two reaction ratios were tested, 1:1 and 1:2.

The undigested gene (containing EcoRI and BamHI restriction sites) concentration was 72 ng/μL. A 1:20 dilution of the gene was prepared (3.6 ng/μL).The supplied TOPO vector concentration was 15 ng/μL.

$$\frac{780\ bp}{5764\ bp} \times 15\ ng\ of\ vector = 2\ ng\ of\ gene \qquad [6]$$

$$\frac{2\ ng}{3.6\ ng/\mu l} = 0.55\ \mu L\ of\ insert \qquad [7]$$

***Table 11**. TOPO® Cloning Reaction mix*

Reagent	Volume (μL)		
	Ratios		
	1:1	1:2	Control
Milli-Q H2O	3.45	2.9	4
Salt Solution	1	1	1
TOPO® Vector	1	1	1
kred with pGEX-4T-1 primers undigested (72.0 ng/μL)	0.55	1.1	-

The reaction tube was mixed gently and incubated at room temperature for 7'.

After the reaction, the One Shot® TOP10 Chemical Transformation Protocol was followed to transform chemically competent TOP10 cells:

1. Add 3 µL of the TOPO Cloning reaction from **Performing the TOPO® Cloning Reaction**, into a vial of One Shot® TOP10 Chemically Competent *E. coli* and mix gently. **Do not mix by pipetting up and down**.
2. Incubate on ice for 30 min.
3. Heat-shock the cells for 30 s at 42°C without shaking.
4. Immediately transfer the tubes to ice.
5. Add 250 µL of room temperature S.O.C. medium.
6. Cap the tube tightly and shake the tube horizontally (200 RPM) at 37°C for 1 h.
7. Spread 100-200 µL from each transformation on a pre-warmed selective plate and incubate overnight at 37°C. We recommend plating two different volumes to ensure that at least one plate will have well-spaced colonies.

For more information about the kit please visit the following link: https://tools.thermofisher.com/content/sfs/manuals/pettopo_man.pdf

1.11 Colony PCR

In order to screen positive clones within a group of transformants that grew on an LB agar plate a colony PCR method was assessed. This method consists in smearing a piece of each colony we pretend to screen and mixing (by pipetting) each piece in a PCR tube containing 10 µL of the prepared PCR reaction mix (Table 12).

***Table 12**. Colony PCR reaction mix*

Reagent	Volume (µL)
Milli-Q H2O	6.8
DreamTaq Green Buffer (10x)	1
dNTP's (2,5 mM)	1
5' primers (10 µM)	0.5
3' primers (10 µM)	0.5
DreamTaqPolimerase (0,5 U/µL)	0.2

The optimized colony PCR conditions for the T7 promoter primers and for the pGEX-4T-1 colony primers are described in Table 13 and in Table 14, respectively.

***Table 13**. PCR conditions for T7 promoter primers*

STEP	TEMPERATURE	TIME	CYCLES
1	95 °C	7 minutes	1
2	95 °C	30 seconds	
3	57 °C	30 seconds	30
4	72 °C	1minute	
5	72 °C	10 minutes	1

***Table 14**. PCR conditions for pGEX-4T-1 colony primers*

STEP	TEMPERATURE	TIME	CYCLES
1	95 °C	7 minutes	1
2	95 °C	30 seconds	
3	67 °C	30 seconds	30
4	72 °C	1 minute	
5	72 °C	10 minutes	1

The following table contains the sequences of the 5' and 3' primers used in colony PCR. The same primers were also used in the sequencing mixes required by the sequencing service provider.

***Table 15**. 5' and 3' primer sequences used in colony PCR*

Primers	Sequence
T7 promoter 5'	TAATACGACTCACTATAGGG
T7 promoter 3'	GCTAGTTATTGCTCAGCGG
pGEX5'	GGGCTGGCAAGCCACGTTTGGTG
pGEX 3'	CCGGGAGCTGCATGTGTCAGAGG

1.12 Agarose gel electrophoresis

All DNA samples to be analysed by gel electrophoresis were loaded into a gel of 0.8% agarose in TAE 1x Buffer containing 0.2 µg/mL of EtBr. TAE 1x was also used as running buffer on the electrophoresis chamber. Electrophoresis runs were made at a 50-100 V voltage using a PowerPac™ Basic Power Supply (Bio-Rad).

Samples containing a dyed buffer such as FastDigest Green Buffer or DreamTaq Green buffer were loaded directly into the gel. From samples that did not contain any dying substance 2.5 µL were mixed with 0.5 µL of 6X LD and loaded into the gel.

1.13 Mackerey-Nagel Plasmid DNA purification (adapted from user's manual)

The NucleoSpin® Plasmid (NoLid) kit was used to extract all plasmids throughout this project. The following protocol was followed to obtain the plasmid DNA from a 10 mL *E.coli* culture. For more information about the kit please visit www.mn-net.com.

Isolation of high-copy plasmid DNA from *E. coli*.

1. Cultivate and harvest bacterial cells
2. Inoculate a 100 mL flask containing 10 mL of LB and incubate overnight at 37 ºC.
3. Prepare two 2 mL eppendorfs. For each eppendorf, use 2 mL of the saturated culture and pellet cells in a standard benchtop microcentrifuge for 30 s at 11,000 x g. Discard the supernatant and remove as much of the liquid as possible. Repeat this process one more time on the same tube where the previous pellet was obtained.
4. Cell lysis

4.1 Add 250 µL Buffer A1. Resuspend the cell pellet completely by vortexing or pipetting up and down. Make sure no cell clumps remain before addition of Buffer A2.

4.2 Add 250 µL Buffer A2. Mix gently by inverting the tube 6–8 times. Do not vortex to avoid shearing of genomic DNA. Incubate at room temperature for up to 5 min or until lysate appears clear.

4.3 Add 300 µL Buffer A3. Mix thoroughly by inverting the tube 6–8 times until blue samples turn colourless completely. Do not vortex to avoid shearing of genomic DNA.

4.4 Centrifuge for 5 min at 11,000 x g to clarificate the lysate

5. Bind DNA

5.1 Place a column in a collection tube (both provided by the kit) and transfer 750 µL of the last supernatant obtained onto the column. Centrifuge for 1 min at 11,000 x g. Discard the flow-through of the collection tube and place the column back into it.

6. Wash the silica membrane

6.1 Add 600 µL of Buffer A4. Centrifuge for 1 min at 11,000 x g. Discard the flow-through of the collection tube and place the column back into it.

6.2 Add 500 µL of Buffer AW. Centrifuge for 1 min at 11,000 x g. Discard the flow-through of the collection tube and place the column back into it.

7. Dry the silica membrane

7.1 Centrifuge for 2 min at 11,000 x g. Discard the collection tube. This step is to eliminate residual ethanol that might inhibit further enzymatic reactions.

8. Elute DNA

8.1 Place the column in a 1.5 mL tube and add 50 µL of buffer AE into the centre of the membrane. Incubate for 1 min at room temperature. Centrifuge for 1 min at 11,000 x g.

1.14 Measurement of the concentration of double-stranded DNA

The concentration of cleaned PCR products (amplified from the genome of Vp) and extracted plasmids was measured by spectrophotometry. Samples were loaded (2 µL) into a Take3™BioTek® micro-volume plate with 16 microspots of 2 µL. The same elution buffer in which the ds-DNA was eluted was used as a blank (2 µL). The Take3™ plate was read in a BioTek® Powerwave™ XS2 at 260 and 280 nm using the standard Gen5 protocol of the Gen5™ software.

1.15 Enzymatic digestion of the extracted plasmids

To make sure the extracted plasmid was worth sending to a sequencing procedure, different enzymatic digestions were made to verify once again if the there was an insert and if it had the correct size. A single digestion with one of the target restriction enzymes and double digestion with both 5' and 3' restriction enzymes were done to obtain respectively the linearized plasmid and the plasmid separated from the insert. Another checking method was to find a restriction enzyme that would only cut inside the gene and not on the plasmid at all. By making a digestion reaction with such an enzyme like Hind III on the recombinant plasmid and comparing the result with the recombinant plasmid itself, it was possible to check if the plasmid linearized or not, giving proof if the gene was or was not inserted into the plasmid.

2. EXPRESSION

2.1 Incubation of expression strain and IPTG Induction

To start the expression study of the KRED a pre-inoculum of one of the expression strains containing one of our cloned vectors was grown overnight. The morning after, the pre-inoculum $OD_{600\ nm}$ was measured using an Ultrospec 1000 (Amersham Pharmacia Biotech Inc.) spectrophotometer. To find the volume of pre-culture necessary to add to a new flask containing 10 or 100 mL of fresh sterilized medium in order to obtain a starting $OD_{600\ nm}$ of approximately 0.1 the following calculus was used:

$$\frac{0.1 \times A}{B} = C \qquad [8]$$

where A is the volume of new medium to inoculate, B the pre-culture $OD_{600\ nm}$, and C the volume of pre-culture necessary to add to the new medium in mL.

The new inoculated flasks were then incubated at 37 ºC 150 RPM. The $OD_{600\ nm}$ was monitored until it reached a value of 0.5-0.6. At this point of time cultures were induced with 1mM IPTG and then incubated at the desired expression temperature. In cases where a heat shock was meant to take place, cultures were only induced after it. To trigger the heat shock, the cell flasks were dipped in an ice bath for 5-10 minutes.
Throughout this project, two different inducing times were tested, 3 hours and 20 hours of induction.

After the pretended induction time, cells were harvested according to the following steps:

Note: Please take into account that at each harvest, 1mL of culture was pelleted at 11,000 x g for 30 seconds on a 1.5 mL Eppendorf for further whole cell analysis by SDS-PAGE.

1. Split the culture evenly in two 50 mL weighted falcon tubes
2. Centrifuge at 9,000 x g for 15 min at 4ºC
3. Discard the supernatant and resuspend each pellet in approximately 20 mL of sodium phosphate buffer pH 7.0
4. Centrifuge at 9,000 x g for 15 min at 4ºC
5. Discard the supernatant and measure the wet weight of the pellet

2.2 Sonication

Pellets were resuspended with a volume in mL of 6 mM imidazole buffer corresponding to twice the wet weight of each pellet in grams. The resuspended samples were sonicated in contact with an ice bath. For each sample, 5 sonication cycles of 30 seconds at 90% power were performed, with one minute interval between cycles.

Samples were then centrifuged for 45 minutes at 15,500 x g and 4 ºC.

The cell debris (sonicated pellet) was separated from the cell crude extract by pipetting.

2.3 Measurement of protein concentration

The protein concentration of the cell extracts was measured with the Bradford method (53). The calibration curve was produced using bovine serum albumin. Optical density was measured at 595 nm and 1 mL of Bradford reagent was used as a blank. Each sample was prepared for reading by mixing 900 µL with 100 µL dH_2O + sample (i.e. 95 µL of dH_2O + 5 µL of sample).

2.4 Sample preparation for SDS-PAGE electrophoresis

Note: Please take into account that all samples were denatured for 10 minutes at 98 ºC in a Thermomixer comfort 1.5 mL (Eppendorf®) before being loaded into the gel.

The procedures to prepare the three different types of samples that were analysed by SDS-PAGE electrophoresis are described below.

2.4.1 Cell Extract

The following calculus was used to load a protein content of 20 µg into the gel:

$$\frac{20\ \mu g}{A} = B \qquad [9]$$

where A corresponds to the protein concentration of the cell extract measured with the Bradford method and B corresponds to the volume necessary to load into the gel.

Each sample was prepared by mixing the cell extract with the DEN+ solution on a 1:1 ratio. The sample volume loaded into the gel was twice the volume B.

In cases were the Bradford signal was too low samples were prepared by mixing 90 µL of sample with 10 µL of concentrated DEN+. Afterwards, 20 µL of the sample were loaded into the gel.

2.4.2 Whole cells

The following calculus was used to concentrate a cell pellet (1 mL culture) taking in consideration an $OD_{600\ nm}$ = 9

$$\frac{A}{9_{OD}} = B \qquad [10]$$

where A corresponds to the $OD_{600\ nm}$ measured at the harvest of the culture and B corresponds to a volume in µL.

Each sample was prepared by adding ½ B of Tris-HCl SDS pH 8.8 and ½B of DEN+ and resuspending the cell pellet. The volume loaded into the gel was 20 µL.

2.4.3 Sonicated pellet

To prepare the sonicated pellets, a small piece of the pellet was smeared with a loop and resuspended in 100 µL of Tris-HCl SDS pH 8.8 + 100 µL of DEN+. The volume loaded into the gel was 20 µL.

2.5 SDS-PAGE: gel preparation, run and staining

2.5.1 SDS-PAGE gel preparation

The protocol used for the polyacrylamide gel preparation is briefly described below:

1. Prepare the resolving and stacking gel solutions without APS or TEMED and as described in Table 16 and 17, respectively.
2. Place the comb into the assembled sandwich plates. With a marker, place a mark on the glass plate 1 cm below the teeth of the comb. This will be the level to which the resolving gel is poured. Remove the comb.
3. Add the APS and TEMED to the resolving gel solution, mix it gently and pour it to the mark.
4. Immediately overlay the resolving gel solution with dH_2O and wait (45 minutes) until it polymerizes.
5. Remove the top layer of dH_2O and dry the area above the separating gel with filter paper before pouring the stacking gel.
6. Add the APS and TEMED to the stacking gel solution and mix it gently. Rinse the cassette completely with the solution and place the comb inside the sandwich. Finally let the gel polymerize (30 min).

The two necessary recipes for casting the polyacrylamide gel are described in the tables below.

***Table 16**. Mix for Resolving gel 11% preparation*

Reagent	Volume
dH_2O	1.84 mL
Resolving buffer pH 8.8	1.13 mL
Acrylamide/Bis-acrylamide, 30%	1.84 mL
APS, 5%	50 µL
TEMED	10 µL

***Table 17**. Mix for Stacking gel 4.5% preparation*

Reagent	Volume
dH_2O	1.55 mL
Stacking buffer pH 6.8	750 µL
Acrylamide/Bis-acrylamide, 30%	450 µL
Blue bromophenol	250 µL
APS, 5%	50 µL
TEMED	20 µL

The gel polymerized inside Bio-Rad Mini-PROTEAN® Spacer Plates with 0.75 mm Integrated Spacers.

The casted gel was embedded in a Bio-Rad Mini-PROTEAN® Tetra System chamber which was filled with PAGE Running Buffer 1x. Throughout the experimental work two different protein markers were used: Pierce™ Prestained Protein MW Marker and SigmaMarker S8445. The electrophoresis was set to run at 10-15 mA with a Amersham Pharmacia Biotech Inc. electrophoresis power supply 601 (*GE Healthcare*).

Once the electrophoresis finished the gel was retrieved and stained with a Comassie Blue solution and then destained using destaining solution.

2.6 Affinity purification assay

Since the protein was produced with an N-terminal six histidine tag an affinity chromatography column was used to attempt to purify the protein. The resin used was a HIS-Select® Nickel Affinity Gel (approximately 5mL) that was packed inside a glass column and preserved in 20% ethanol.

To equilibrate the column the resin was washed with 6mM Imidazole buffer (approximately 30mL). Then the crude extract that was previously dissolved in 6mM imidazole buffer (approximately 3 mL), filtered with a Minisart® NML syringe filter(16534-K) of 0.2 µm pore size and loaded into the column. At this point of time the first sample was recovered and named "flow-through". Once again the resin was washed with 6mM Imidazole buffer (approximately 30mL). The elution of the protein with the His-tag was done by washing the resin with 250 mM imidazole buffer. The first 8 fractions of 1 mL eluted were recovered and the resin kept being washed till a volume of about 30 mL was spent. Finally the column was re-equilibrated with 20% ethanol.

2.7 Solubilisation of inclusion bodies (IB's), denaturing, and protein refolding

Since the protein was mainly overexpressed as IB's, a simple protein solubilisation protocol was followed with some modifications (54)(55)(6). This protocol is described below:

1. Centrifuge the cell lysate at 4°C for 40 min at 15,000 x g.
2. Resuspend the pellet in 10 mL wash buffer containing 1% Triton X-100 and 1 M urea per gram cell wet weight. Incubate at room temperature for 5 min.
3. Centrifuge the cell lysate at 4°C for 40 min at 15,000 x g.
4. Repeat step 3 twice.
5. Resuspend the pellet in 10 mL wash buffer and centrifuge the suspension at 4°C for 30 min at 235,000 x g.
6. Resuspend the inclusion bodies in extraction buffer. Use enough buffer so that the final protein concentration in the solution is 1 mg/mL. Incubate the solution at room temperature for 60 min.
7. Dilute the solution tenfold with extraction buffer (final protein concentration is 0.1 mg/ml) and dialyze overnight against a 100 fold volume of wash buffer.
8. Centrifuge the dialysate at 4°C for 30 min at 235,000 x g.

Chapter 3

Results and discussion

1. CLONING OF THE KRED GENE IN *E. coli*

The sequence of kred is presented in Table 18. Three different attempts of cloning the gene in three different vectors were made. Planning of recombinant plasmids construction, containing the gene in frame with an N-terminal affinity tag, was done using SnapGene® software. The constructs are presented in figures 10, 11 and 12.

***Table 18**. kred gene sequence*

Gene	Sequence
kred	Atggattatgcaattgttacaggagcttcacaaggattgggaaaagaaatagctaagttgctattgagttatggagtatat gtaatcggtatctcccgaagtaagattactgatttacacagtgtagctgataaatataaggttacctttcaaagtattgctt atgatttagtggatgtagcgaatctagaagttcttcttgatacaatcactacaaagcatatgaaaaatgacgttagaaag ctttatcttgtgaataatgcagctgttattgaaccgattaatcagggaactcatatattacctacgtctttaaattaccattt tcagattaatgctgtcgcgccaatggcattaacgaatttgttcttacagtttgcgaatcaaaaacaaattccactaacttg tttaacagtgacttctggtgctgctgagaaaccgacgtttggatggactgcttattgtagttctaaggctagtgtaaacatg tatacaaaaaccattgctaaagagcaagatgagctaaatacgaaacatactatcattgctttcagtccaggtgttatgga tacggagatgcaagcacaaattcgttccagtgaaaaatctgcatttcaagcagtagacacatttaaagcgtactatgaaa ataaacggttgcaacatgcagaaaaaatcgcctctatagctacagagatcatgttagacaataacatagaaaatggacg aatatattatgcacgtgattatttatcttaa

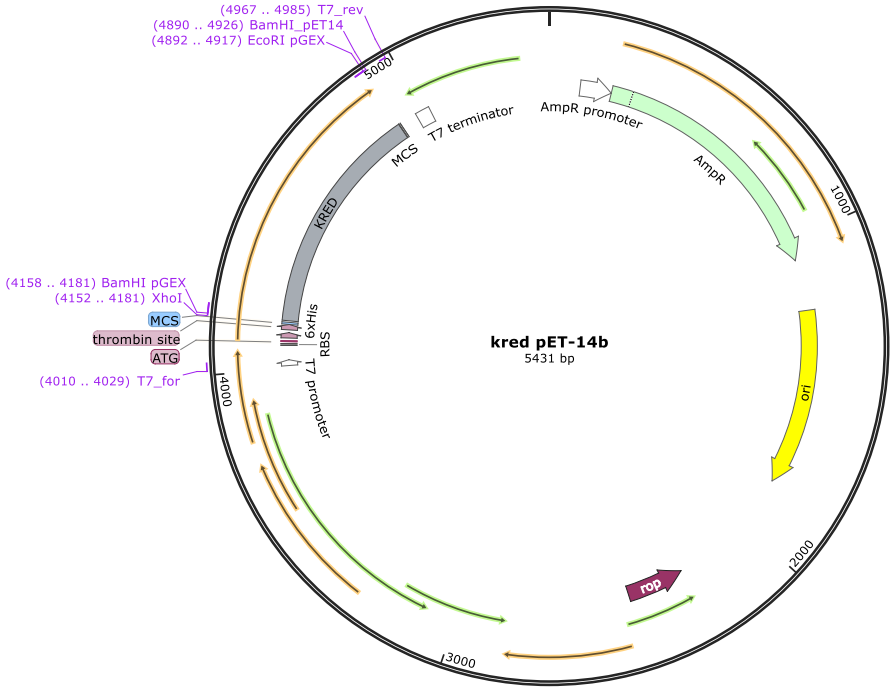


Figure 10. kred pET-14b construct. Image obtained using SnapGene® software

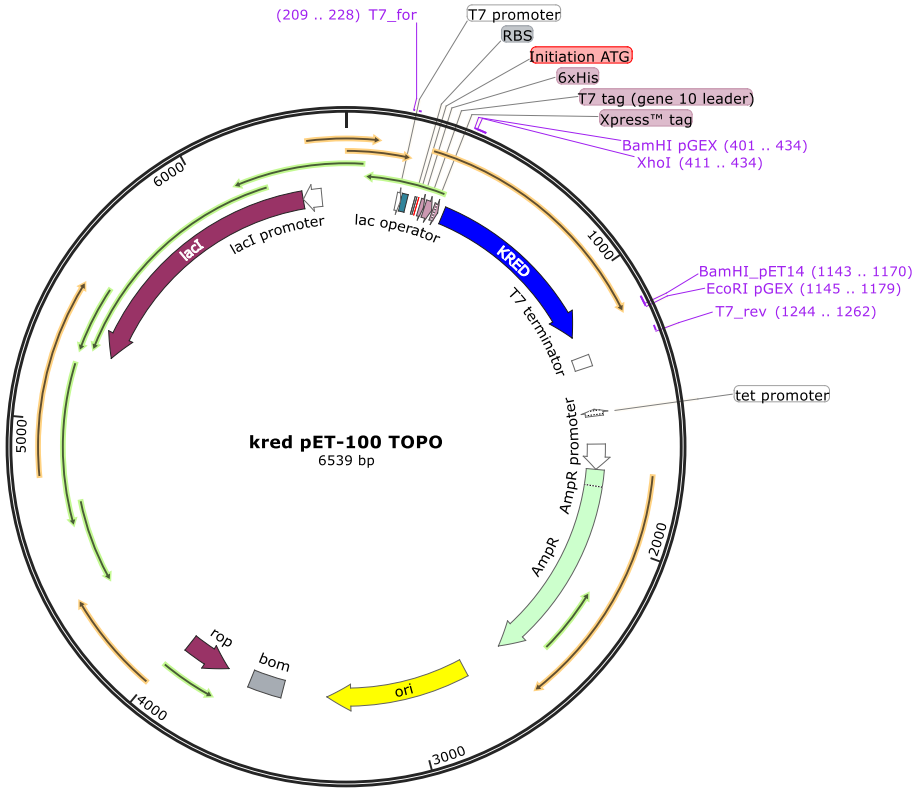


Figure 11. kred pET-100 TOPO construct. Image obtained using SnapGene® software

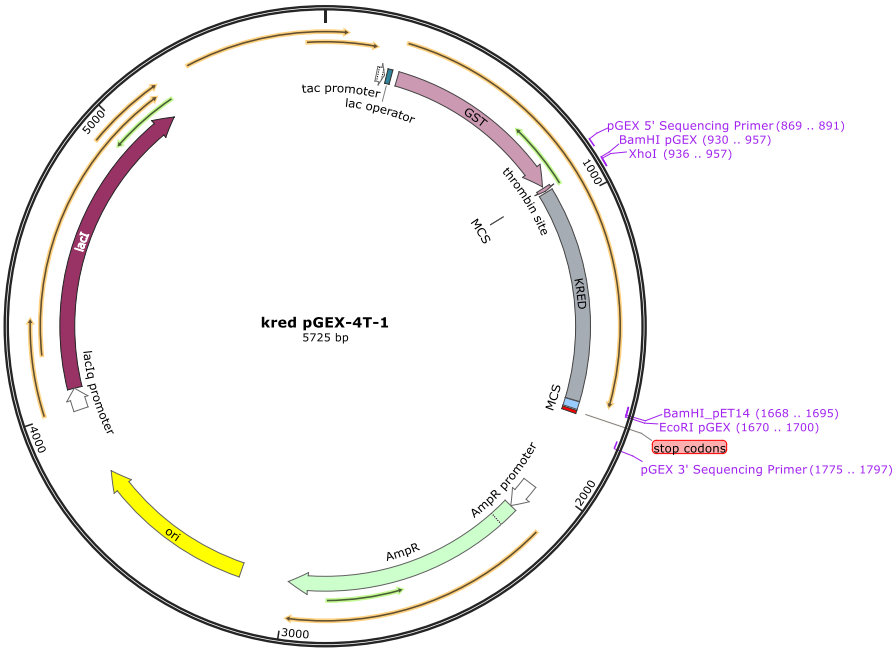


Figure 12. kred pGEX-4T-1 construct. Image obtained using SnapGene® software

1.1 Amplification of kred from the genome of Vp

In order to find out the best melting temperature (Tm) for each set of primers, a PCR was conducted as described in chapter 2, subchapter 1.2. The gene had a length of 759 bp.

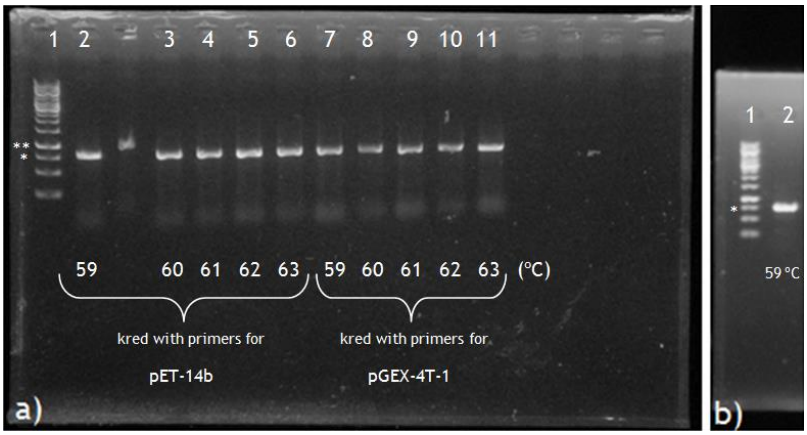


***Figure 13. a)** and **b)** Amplification of kred from the genome of Vp. **a)** Lane 1 - 1 kb marker; lanes 2 to 6 - Tm gradient from 59 to 63 ºC with primers for pET-14b; lanes 7 to 11 - Tm gradient from 59 to 63 ºC with primers for pGEX-4T-1; **b)** Amplification of kred with primers for pGEX-4T-1 with Tm= 59 ºC. Marker description: * - 750 bp; ** - 1000 bp*

Once amplified with the selected primers, the PCR product should have about 20 bp more than the actual gene size. Therefore it is expected that a correct amplification should appear slightly above the 750 bp marker band. For this reason it is possible to conclude that 59 ºC was the best annealing temperature for both amplifications (Figure 13.a)).

To obtain more PCR product, the reaction was repeated but this time with only one annealing temperature, 59ºC. To demonstrate the result of this amplification, Figure 13.b) shows on lane 2 the amplification of kred with BamHI_pGEX-4T-1_FW and EcoRI_pGEX-4T-1_RV primers.

All PCR products were then cleaned and digested according to subchapters 1.3 and 1.4.3, respectively, of chapter 2. These digested products were not analysed by electrophoresis since it was impossible to distinguish them from the amplified kred. (Figure 13.b)).

1.2 Double digestion and dephosphorylation of the pet-14b and pGEX-4T-1.

Vector preparation for the cloning procedure was done by digesting each one with the chosen restriction enzymes and dephosphorylating them according to the sub-chapters 1.4.1, 1.4.2 and 1.5, of chapter 2. Figure 14 presents the gel containing the non-digested and the double digested vectors.

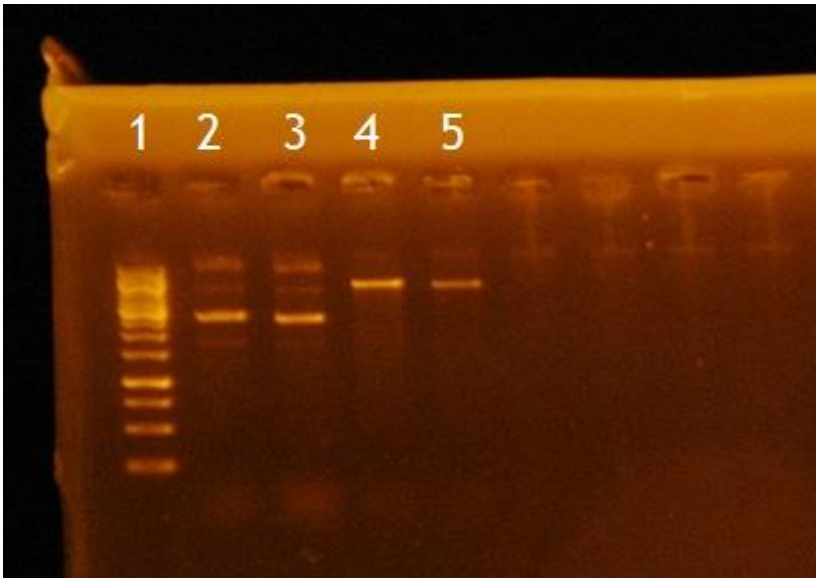


Figure 14. *Enzymatic digestion of pGEX-4T-1 and pET-14b. Lane 1 - 1 kb marker; lane 2 - undigested pGEX-4T-1; lane 3 - undigested pET-14b; lane 4 - pGEX-4T-1 digested with BamHI and EcoRI; lane 5 - pET-14b digested with XhoI and BamHI.*

Each well was loaded with 2.5 µL of sample + 0.5 µL of 6x Loading Buffer since the normal (non-dyed) FD Buffer 10x was used for this digestion. The non-digested pGEX-4T-1 and the non-digested pET-14b are represented in lanes 2 and 3, respectively. On these two lanes, from top to bottom it is possible to see the relaxed, nicked and supercoiled forms of the plasmids (56). The digested pGEX-4T-1 and pET-14b are represented in lanes 4 and 5, respectively. As expected, the linearized pGEX-4T-1 position was a slightly higher compared to that of pET-14b.

1.3. Cloning kred in pET-14b

A ligation reaction between the kred and the digested pET-14b was performed according to chapter 2, sub-chapter 1.6. Afterwards, DH5α™ cells were transformed with the ligation mix according to chapter 2, sub-chapter 1.7. Screening for positive clones was done by colony PCR. The following gel (Figure 15) shows a screening of 27 colonies.

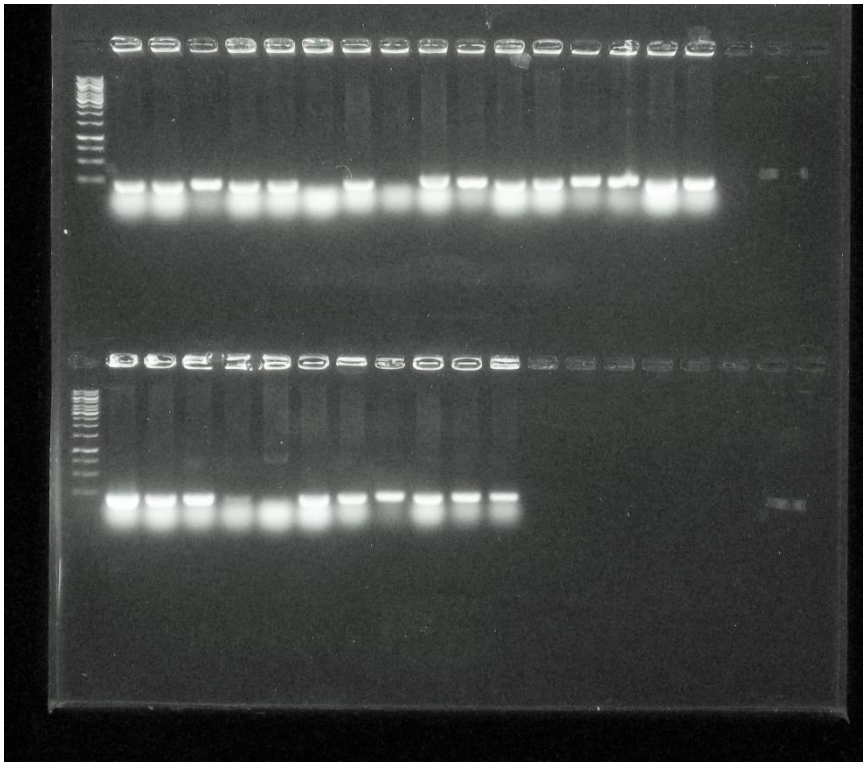


***Figure 15**. Colony PCR of DH5α™ cells transformed with kred pET-14b ligation product. All the 27 colonies were not successfully cloned*

To better evaluate the previous gel results, the PCR product size with and without the insert was measured using SnapGene® software. If a positive clone would be screened, a product with 976 bp was expected to be seen in the gel. In case the screened clone did not have the insert, an amplification of 216 bp should appear. The measurements can be seen in Annex 2, pictures A and B. There is no amplified band in the 1000 bp region, meaning that there is no positive clone.

1.4. Cloning kred with Champion™ pET Directional TOPO® Expression Kit

A colony PCR (Figure 16.a)) was performed after cloning the kred with the help of a cloning kit as described in sub-chapter 1.8 of chapter 2.

The expected size of a positive clone PCR product is 1054 bp. On the other hand, the expected size of a negative clone PCR product is 279 bp. This information was obtained using SnapGene® software and can be seen in Annex 2, pictures C and D.

A total of 14 colonies were screened and 13 of them were positively cloned as seen in Figure 16 a).

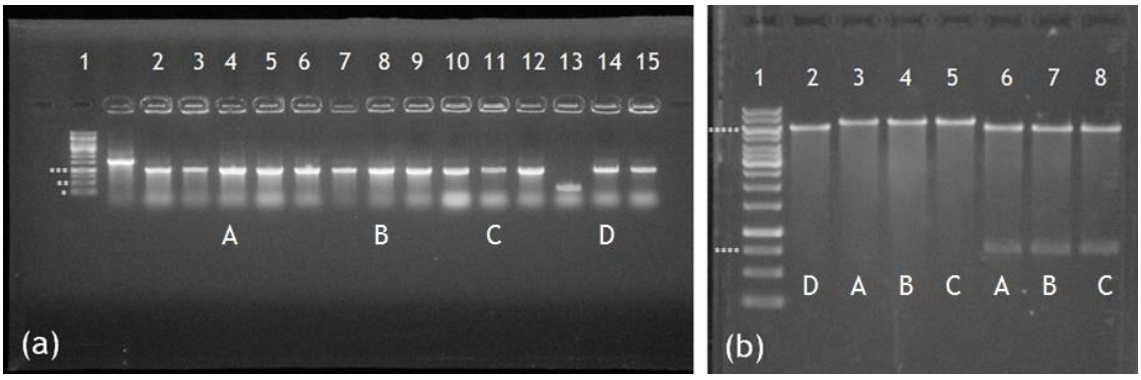


Figure 16**. **(a)** Colony PCR of TOP10 cells transformed with the cloning kit product (kred pET-100 TOPO). Lane 1 - marker 1 kb; All lanes from 2 to 15 present a positive insert amplification, except lane 13; **(b)** Enzymatic digestion of plasmids extracted from the same colonies presented in lanes 4, 8, 11 and 13 (A, B, C and D, respectively). Lane 1 - marker 1 kb; Lane 2 - Single digestion of plasmid D with BamHI; Lanes 3, 4 and 5 - Single digestion of plasmids A, B and C with BamHI; Lanes 6, 7 and 8 - Double digestion of plasmids A, B and C with BamHI and EcoRI. Marker description: * - 250 bp; ** - 500 bp; *** - 1000 bp; * - 750 bp; ***** - 6000 bp.*

Four colonies named A, B, C and D (from lanes 4, 8, 11 and 13, respectively)were saved for further work by re-picking and inoculating them in fresh LB agar plates containing 100 µg/mL of ampicillin and incubating at 37 ºC overnight.
Plasmids of the mentioned colonies were extracted according to sub-chapter 1.11 of chapter 2.

To further confirm the last result presented, the extracted plasmids were digested to verify the presence of the insert. On lanes 6, 7 and 8 two products can be seen, the double digested plasmid and the correct insert near the marker band of 750 bp. The positive cloning can also be confirmed by comparing the single digested plasmids of lanes 3, 4 and 5 with the double digested ones. There is a clear difference in size since the single digested plasmids contain the insert. A third confirmation can be observed, since the negative plasmid of lane 2 that does not contain the insert has the same size of the plasmids of lanes 6, 7 and 8.

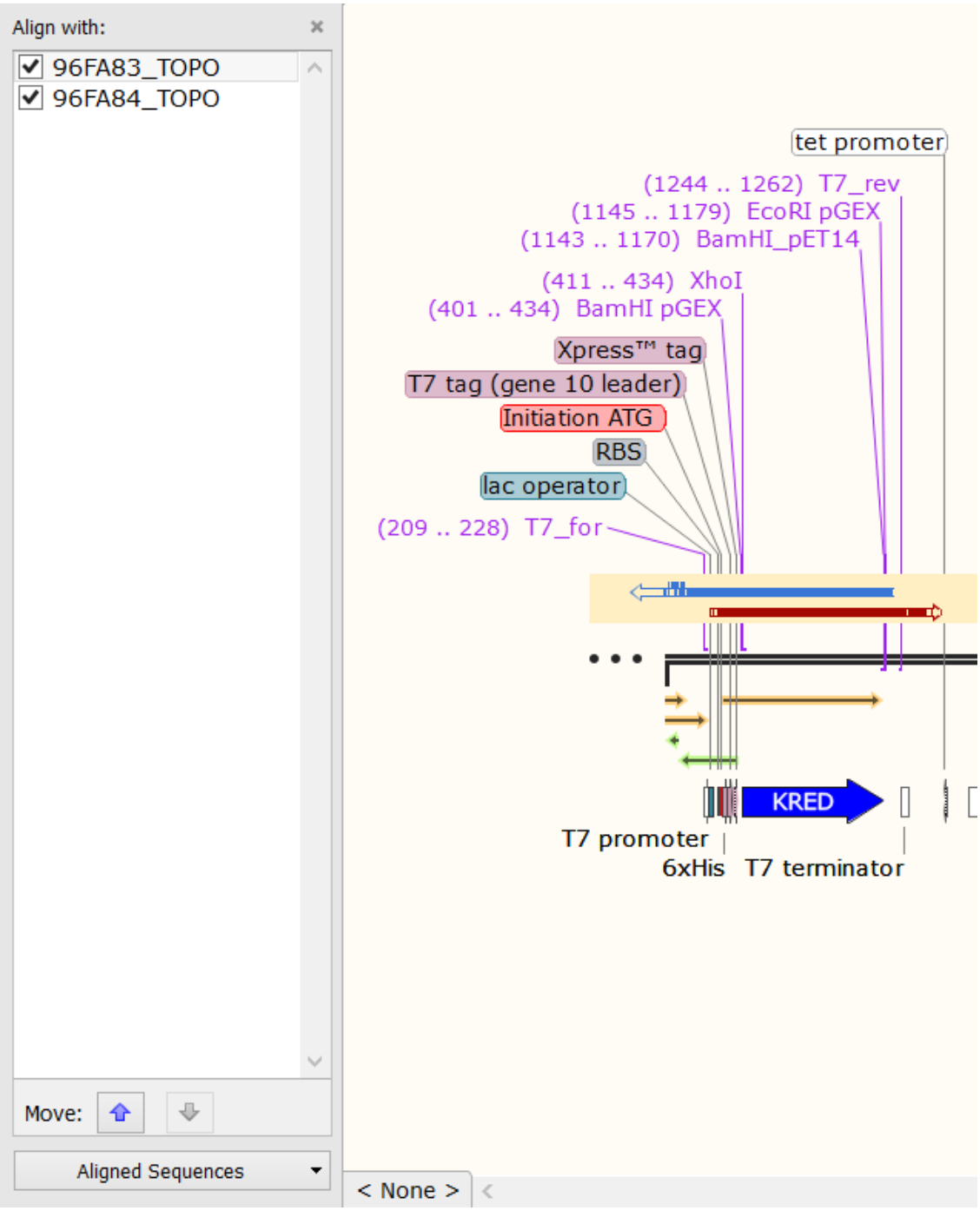


***Figure 17**. Sequence alignment of the forward and reverse sequencing results with the kred pET-100 TOPO vector. No mutations were verified within the kred range. Image obtained using SnapGene® software.*

Finally, regarding the positive outcome, the cloned plasmid was sent for sequencing to ensure that there were no mutations in the gene sequence. The insert was sequenced by GATC Biotech AG by LIGHTRUN™ sequencing. The plasmid as well as the colony PCR primers (FW and RV) for the insert region were sent to the company according to their instructions. Two separated files were received from the sequencing service provider. Each file contained a chromatogram generated by the sequencing process with each primer, the forward and the

reverse. The tool "Align multiple sequences" of SnapGene® software was utilized to align both sequences with the original construct file. The program highlighted each mutated base pair or sequencing gap. Figure 17 reveals that there was no mutation in the gene sequence, confirming that the cloning process was successful.

1.5. Cloning of kred pGEX-4T-1

Since the colony PCR primers used in the pET-14b and pET-100 TOPO cloning were commonly used (T7 promoter), their best performing conditions, especially the Tm, were already known. Nevertheless, to find out the best suitable annealing temperature of the primers designed for the pGEX-4T-1 vector a PCR between the primers and the empty vector was performed according to the tables 11 and 13 of the sub-chapter 1.9, chapter 2. Nonetheless a gradient of melting temperatures was tested, from 63 ºC to 67 ºC.

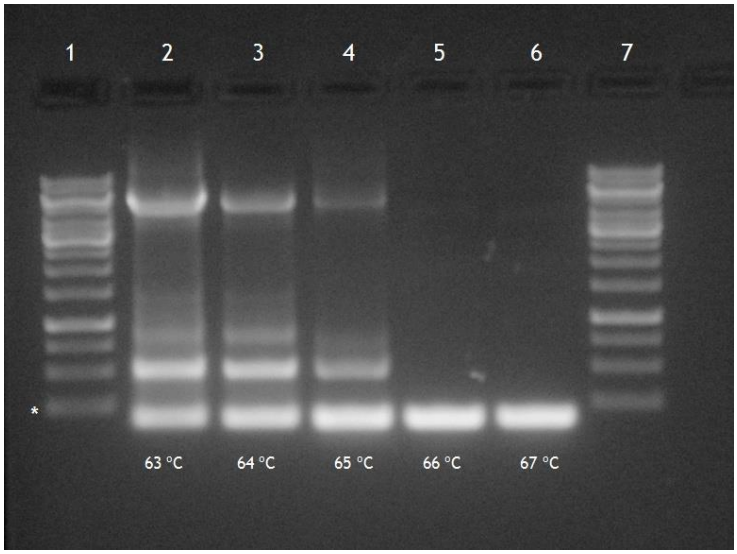


***Figure 18**.PCR using the pGEX-4T-1 vector and the pGEX 5' and pGEX 3' primers to find the most suitable Tm for colony PCR reaction. Lanes 1 and 7 - Marker 1kb; Lane 2 - Tm= 63 ºC;Lane 3 - Tm= 64 ºC; Lane 4 - Tm= 65 ºC; Lane 5 - Tm= 66 ºC; Lane 6 - Tm= 67 ºC. Marker description: * - 250 bp.*

As it can be seen on Figure 18, at 63 ºC, 64 ºC and 65 ºC there were several unspecific amplifications. Only at 66 ºC and 67 ºC there was a single amplification that was below the 250 bp marker. This was expected since the frame between the forward and the reverse primers had a length of 173 bp.

The kred was cloned in the pGEX-4T1 according to the procedures from sub-chapter 1.2 till 1.9 of chapter 2. A colony PCR reaction was performed with the optimized annealing temperature to search for any positive clones. (Figure 19 a)).

In order to analyse the results of the gel presented in Figure 16, the PCR product size with and without the insert was measured using SnapGene® software. A positive clone would yield a product with 929 bp. In case the screened clone did not have the insert, an amplification of 173 bp should appear. The measurements can be seen in Annex 2, pictures E and F.

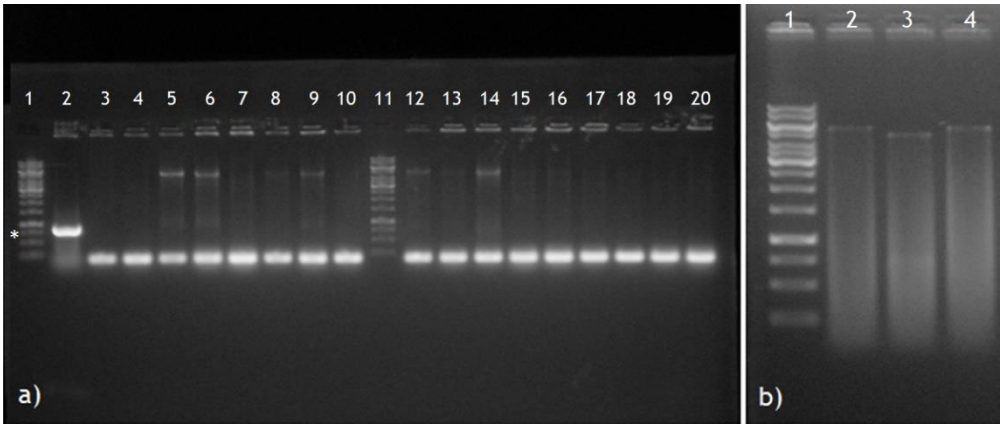


***Figure 19. a)** Colony PCR of DH5a™ cells transformed with kred pGEX-4T-1 ligation product. Lanes 1 and 11 - marker 1 kb; Lanes 3 to 10 and 12 to 18 - colonies that were not successfully cloned; Lane 2 - colony positively cloned; Lanes 19 and 20 - negative control colonies of DH5a™ cells transformed with pGEX-4T-1; **b)** Digestion of plasmid extracted from the colony screened in lane 2 of Figure 16.a). Lane 1- Marker 1 kb; Lane 2- Single digestion with BamHI; Lane 3- Double digestion with BamHI and EcoRI; Lane 4- Single digestion with HindIII. Marker description: * - 750 bp.a) and b)*

Lanes 2-10 and 12-18 (Figure 19.a)) are clones that were transformed with the positive construct. Lanes 19 and 20 are two negative controls of clones that were transformed with the native pGEX-4T-1 vector. It is clear that there is a positive clone on lane 2 since the band is slightly above the 750 bp marker band. The colony from lane 2 was re-picked, inoculated in a fresh LB agar plate containing 100 µg/mL of ampicillin and incubated at 37ºC overnight in order to save the clone.

Plasmids of the new positive clones were extracted according to the sub-chapter 1.11 of chapter 2.

To further verify the correct cloning of the gene the extracted plasmid was digested according to the sub-chapter 1.13 of chapter 2. The gel showing the digestion procedure is presented in Figure 19.b). Although it is not possible to see the insert fragment on lane 3 near the 1000 bp zone, there is a difference in plasmid size between the single digestion on lane 2 and the double digestion on lane 3. We can also see that on lane 4 the plasmid got linearized with the Hind III digestion. This enzyme is only capable of cutting inside the gene and nowhere else in the plasmid. The linearized product is also of similar size as the product of lane 2. Both aspects suggest that the gene is present in the plasmid.

The cloned plasmid was sent for sequencing to ensure that there was not any mutation in the gene sequence. The insert was sequenced by GATC Biotech AG by LIGHTRUN™ sequencing. The plasmid as well as the colony PCR primers (FW and RV) were sent to the company according to their instructions. Two separated files were received from the sequencing service provider. Each file contained a chromatogram generated by the sequencing process with each primer, the forward and the reverse.

The tool "Align multiple sequences" of SnapGene® software was utilized to align both sequences with the original construct file. The program highlighted each mutated base pair or sequencing gap. Figure 20 reveals that there was no mutation in the gene sequence, demonstrating that the cloning process was successful.

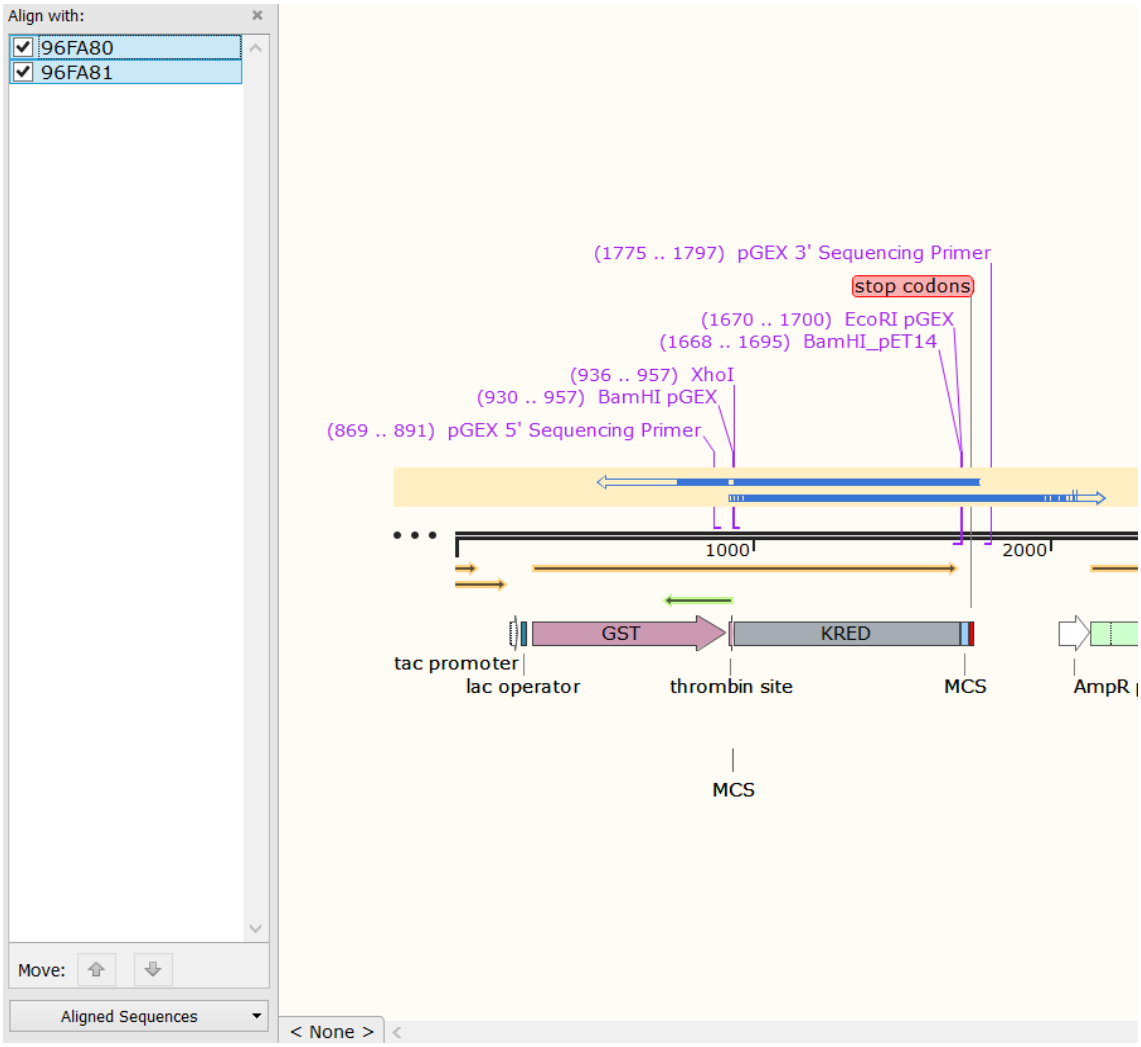


Figure 20. *Sequence alignment of the forward and reverse sequencing results with the kredpGEX-4T-1 vector. No mutations were verified within the kred range. Image obtained using SnapGene® software.*

2. EXPRESSION

Regarding the cloning results shown in sub-chapter 1 of chapter 3, different attempts were made to express KRED with an N-terminal His-tag and GST-tag. In order to estimate the molecular weight size of KRED with the two different tags the SnapGene® software was used. Since the GST fusion protein has a considerable size, its molecular weight was also measured due to the fact that it could be expressed in a negative control.

***Table 19**. Molecular weight in kDa of the different proteins and N-terminal tags (Annex 2, pictures G, H and I). Note that according to the TOPO cloning procedure this His-tag does not contain only six histidines but also some other specific sites such as the entero-kinase site.*

Protein	**Molecular weight (kDa)**
KRED	28.1
His-tag *	4.3
GST-tag	27.9
His-tag KRED	32.4
GST-tag KRED	54.4

As a first expression study, BL21 (DE3) STAR™ cells were transformed with the kred pET-100 TOPO construct. The expression study was done as described in the chapter 2, subchapter 2. Briefly, the culture was centrifuged and its pellet ressuspended in Imidazole Buffer 6 mM. This suspension was then sonicated and centrifuged, yielding a **"sonicated pellet"** (pellet) and a **"cell extract"** (supernatant) that corresponded to the insoluble and soluble fractions, respectively. In parallel, a sample of 1 mL of culture was centrifuged and the pellet was saved (identified as **"whole cells"**). No overexpressed protein was obtained. The gel is presented and described in Annex 3, Picture J.

Despite the negative result, in an effort to obtain the target protein, the same expression study was repeated using an auto-induction medium (A.I) that did not require the use of IPTG as an expression inducer since it contained α-lactose. Once again, as it can be observed in Annex 3, Picture K, no overexpressed protein was obtained.

Considering this work scenario the expression strain was changed. The BL21 CODON+ strain was transformed with the kred pET-100 TOPO construct. The gel presented in Figure 21 reveals the result of an expression study done at 37 ºC according to the subchapter 2 of chapter 2.

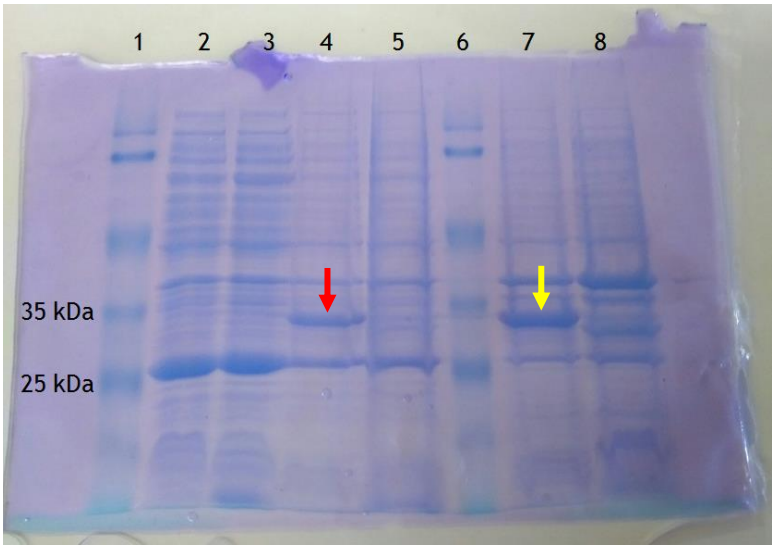


***Figure 21**. SDS-PAGE of expression study of kred pET-100 TOPO with BL21 CODON+, at 37ºC. Lanes 1 and 6 - Pierce™ Prestained Protein MW Marker; Lane 2 - Cell extract kred pET-100 TOPO, 20h IPTG induction; Lane 3 - Cell extract pET-100 TOPO, 20h IPTG induction (negative control); Lane 4- Sonicated pellet kred pET-100 TOPO, 20h IPTG induction; Lane 5- Sonicated pellet pET-100 TOPO, 20h IPTG induction; Lane 7- Whole cells kred pET-100 TOPO, 20h IPTG induction; Lane 8- Whole cells pET-100 TOPO, 20h IPTG induction. Red and yellow arrows indicate the location of the KRED band.*

Lanes 4 and 7 reveal an overexpressed protein. Meanwhile, on the respective negative controls the same protein band is not present, suggesting that the band corresponds to the KRED. Nevertheless on lane 2 there is no overexpressed protein at the same band size level which means that KRED was mainly expressed in an insoluble state and/or as inclusion bodies.

The BL21 CODON+ was also transformed with the kred pGEX-4T-1 construct. The expression of this transformed strain was done at 25 ºC with cold-shock. The cold-shock process aims to favour the production of soluble KRED. This method is based in different facts that lower temperatures may enhance. An immediate consequence of temperature reduction is the elimination of some heat shock proteases that are induced under typical overexpression conditions. (56) Protein hydrophobic interactions may increase with high temperatures, enhancing protein aggregation and the formation of inclusion bodies (57). Moreover *E. coli* chaperone expression is increased at temperatures around 30 ºC (58) (59).Nevertheless, lowering the incubation temperature or performing such shock inhibits replication, transcription and translation (60). Decreased temperatures also affect the efficiency of promoters used in recombinant vectors. Decreased promoter efficiency lowers the protein yield but at the same time favours protein folding. The same transcriptional effect can be obtained when weak promoters are used or when strong promoters are slightly induced (61) (62).The SDS-PAGE of the samples related to the expression study of KRED in BL21 CODON+ at 25ºC with cold-shock is presented in Figure 22.

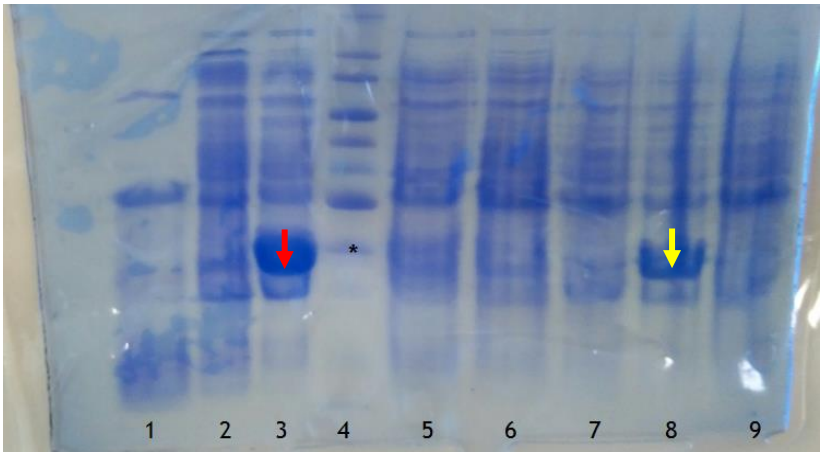


***Figure 22**. SDS-PAGE of expression study of kred pGEX-4T-1 with BL21 CODON+, at 25ºC with cold-shock. Lane 1 - Cell extract kred pGEX-4T-1, 3h IPTG induction, loaded with DEN+ Concentrated; Lane 2 - Cell extract kred pGEX-4T-1, 20h IPTG induction; Lane 3 - Cell extract pGEX-4T-1, 20h IPTG induction; Lane 4- SigmaMarker S8445, * - 29 kDa; Lane 5- Whole cells kred pGEX-4T-1, 3h IPTG induction; Lane 6- Whole cells kred pGEX-4T-1, 20h IPTG induction; Lane 7 - Whole cells kred pGEX-4T-1, non-induced; Lane 8- Whole cells pGEX-4T-1, 20h IPTG induction; Lane 9 - Whole cells pGEX-4T-1, non-induced. The red and the yellow arrow indicate the overexpressed GST fusion protein.*

The gel reveals that only the GST tag in the negative control was overexpressed (lanes 3 and 8). There is no evident band that corresponds to KRED.

Since in the previously reported expression studies no overexpressed soluble protein was obtained, the expression approach was re-planned. The approach followed was adapted from Sandra Harper and David W. Speicher work (63). A total of 6 new colonies were screened from each type of transformants, the ones with kred pET-100 TOPO (CODON+) and the ones with kred pGEX-4T-1 (CODON+), to find the one that produced KRED with the highest yield. Only whole cell samples were analysed.

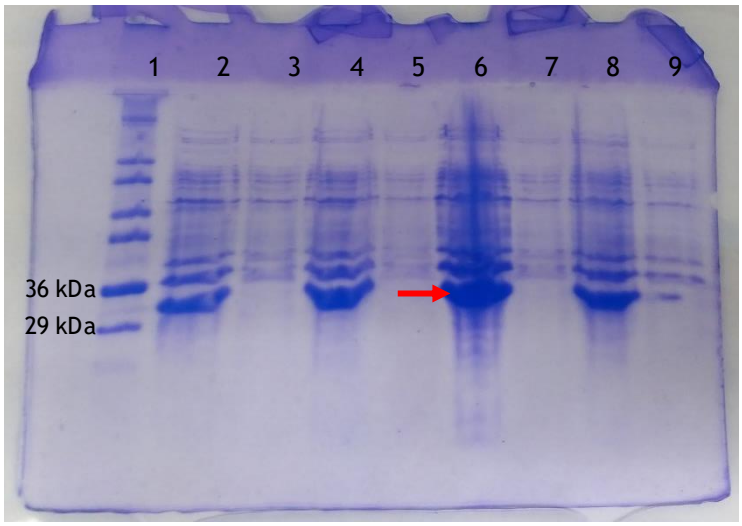


***Figure 23**. Screening of kred pET-100 TOPO (CODON+) transformants. This gel presents 4 out of the 6 screened colonies. Lane 1 - SigmaMarker S8445; Lanes 2 & 3, 4 & 5, 6 & 7, 8 & 9 – the first lane of each pair corresponds to an induced culture, while the second one corresponds to a non-induced culture.*

Regarding the kred pET-100 TOPO (CODON+) transformants, the SDS-PAGE of 4 out of the 6 screened colonies is presented in Figure 23 (lanes 2 & 3, 4 & 5, 6 & 7, 8 & 9). The colony represented in lanes 6(induced) and 7(non-induced) revealed the strongest band of target protein and was therefore selected as the best producer. The screening of the kred pGEX-4T-1 (CODON+) transformants gave an incomprehensible result since all the induced transformants expressed a protein with the same approximate size as the GST fusion protein. This suggests that the transformants were not cloned with the positive construct.

The best His-tag KRED producer chosen from the screening process was submitted to a new expression study. The expression was done both at 25 ºC with heat-shock and at 37 ºC. Initially, only the whole cell samples were run through electrophoresis (Figure 24).

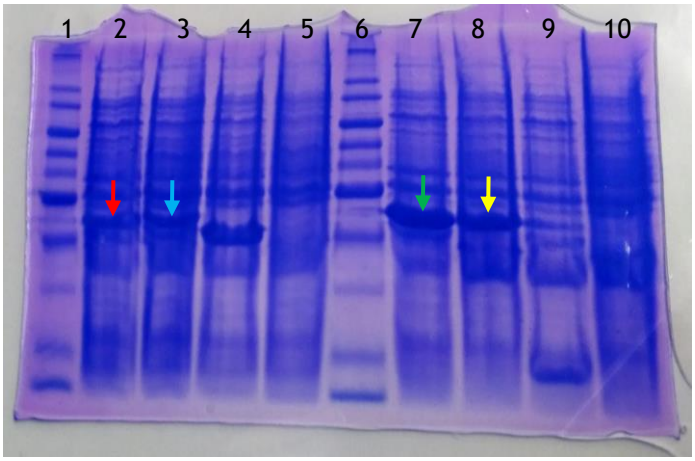


***Figure 24**. Expression study of His-tag KRED best producer (BL21 CODON+) at 25ºC with heat-shock and at 37ºC. Lanes 1 and 6 - SigmaMarker S8445; Lane 2 - Whole cells kred pET-100 TOPO, 25ºC, 3h IPTG induction; Lane 3 - Whole cells kred pET-100 TOPO, 25ºC, 20h IPTG induction; Lane 4 - Whole cells pET-100 TOPO, 25ºC, 20h IPTG induction; Lane 5 - Whole cells kred pET-100 TOPO, 25ºC, non-induced; Lane 7 - Whole cells kred pET-100 TOPO, 37ºC, 3h IPTG induction; Lane 8 - Whole cells kred pET-100 TOPO, 37ºC, 20h IPTG induction; Lane 9 - Whole cells pET-100 TOPO, 37ºC, 20h IPTG induction; Lane 10 - Whole cells kred pET-100 TOPO, 25ºC, non-induced. Coloured arrows indicate the location of KRED.*

In an attempt to purify a small amount of soluble protein, the cell extract of the culture represented in lane 3 (kred pET-100 TOPO 20h induction, 25ºC) was obtained and submitted to an affinity purification assay as described in sub-chapter 2.6. The eluted fractions of the assay were analysed through electrophoresis. This experiment was unsuccessful and no purified soluble protein was observed in the SDS-PAGE. The gels of this purification assay can be seen in pictures M and N of annex 3.

In a last attempt to obtain a soluble protein a solubilisation protocol was performed as described in sub-chapter 2.7. The culture pellets from which the inclusion bodies were retrieved were the kred pET-100 TOPO of 3h and 20h induction at 37 ºC. The whole cell samples that reveal the previously mentioned inclusion bodies can be observed in lanes 7 and 8 of Figure 24. Briefly, the pellets were ressuspended in Imidazole Buffer 6 mM and mixed together. This solution was then sonicated and centrifuged, yielding a **sonicated pellet** (pellet) and a **cell extract** (supernatant) that corresponded to the insoluble and soluble fractions, respectively. The inclusion bodies present in the sonicated pellet were denatured and ressuspended in the extraction buffer. Afterwards the solution with the denatured inclusion bodies was dialysed and centrifuged yielding a supernatant and a pellet that corresponded to the soluble and insoluble fractions, respectively. The SDS-PAGE analysis of the dyalisate can be observed in Figure 25.

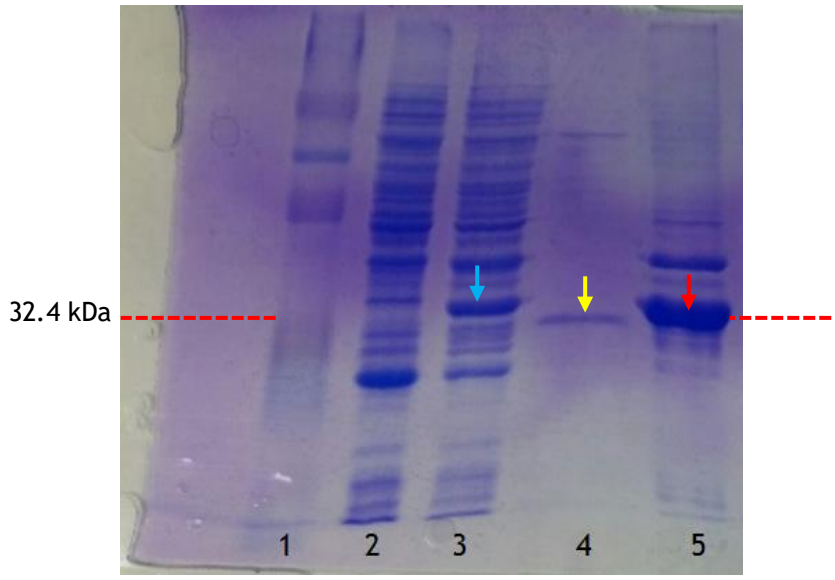


***Figure 25**. SDS-PAGE of KRED inclusion bodies solubilisation, expressed with BL21 CODON+ at 37 ºC. Lane 1 - SigmaMarker S8445; Lane 2 - Cell extract of kred pET-100 TOPO, 3h + 20h IPTG induction; Lane 3 - Sonicated pellet kred pET-100 TOPO, 20h IPTG induction at 25 ºC with cold-shock (not related with the title of this figure, blue arrow indicates the location of KRED); Lane 4 - Soluble fraction of the dialysate; Lane 5 - Insoluble fraction of the dialysate. Yellow and red arrows indicate the location of KRED*

A weak band is perceptible on lane 4 (soluble fraction) that has, approximately, the same molecular weight as the inclusion bodies present in lane 5 (insoluble fraction). This result may suggest that the protein was produced in the soluble state although the band seems slightly lower than the inclusion bodies band. However, the folding state and the amount of loaded protein on each well may have affected the migration of the proteins and the apparent size difference may therefore be an artifact.

The procedures done after screening the best His-tag KRED producer are summarized in flowchart presented in Figure 26.

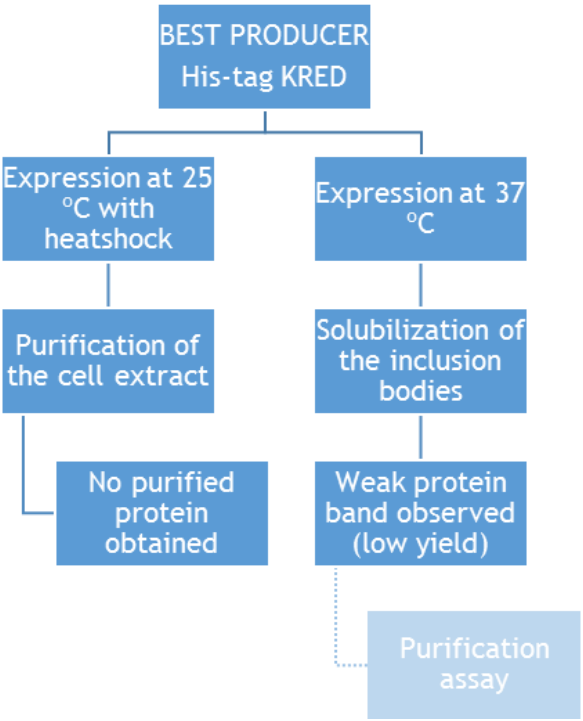


***Figure 26**. Flowchart of the procedures done after screening the best His-tag KRED producer.*

The majority of PTMs (post-translational modifications) present in eukaryotic proteins do not occur in *E. coli* recombinant proteins. These PTMs, such as glycosylation, disulphide bond formation and phosphorylation, are essential in correct protein folding, processing, stability and biological activity. Furthermore, such PTMs might be responsible for increasing the serum half-life of a product. Therefore, recombinant proteins produced with *E. coli* systems just as KRED was, might be insoluble, unstable or inactive (64) (65).

On the other hand, natural protein characteristics might be responsible for their own insolubility. Therefore, some of KRED and His-tag KRED aspects were analysed using online bioinformatics tools. The same features were also evaluated on another short-chain ketoreductase (KRED1-Pglu) belonging to a marine microorganism *Pichia glucozyma* CBS 5766 for comparison. The research group that hosted me during this master's thesis work was able to overexpress KRED1-Pglu in a soluble state in similar expression conditions (66).

Initially the percentage of major hydrophobic residues was determined for each of the three proteins, KRED, His-tag KRED and KRED1-Pglu. To do so, each protein sequence was uploaded in the online tool SAPS (Statistical Analysis of Protein Sequences) of the European Bioinformatics Institute (67)and the LVIFM index (major hydrophobic residues, %) of each one was computed.

***Table 20**. Percentage of major hydrophobic residues*

Protein	Major hydrophobic residues (%)
KRED	29.4
His-tag KRED	27.6
KRED1-Pglu	30.2

There are no major differences between the percentage of hydrophobic residues in KRED1-Pglu and His-tag KRED. Hydrophobicity is assumed as one of the major factors contributing to insolubility events and formation of inclusion bodies. Nevertheless, KRED1-Pglu has a hydrophobicity index higher than His-tag KRED and was soluble upon overexpression while His-tag KRED was not. It can also be observed that the His-tag did not increase the hydrophobicity index.

A secondary structure consensus prediction was also made for KRED and His-tag KRED using the MLRC, DSC and PHD methods available at the Prabi-Gerland Rhone-Alpes Bioinformatic Pole Gerland website (68). Although simulation has shown that the His-tag did not increase the hydrophobicity index of the recombinant protein, the secondary structure prediction aims to discover to what level did the insertion of this affinity tag may have changed the protein secondary structure (Figure 27) and consequently its folding and solubility. The addition of the His-tag caused structural changes such as the decrease of alpha-helix structures (marked as blue) and the increase of extended strands (marked as red) in certain amino acids comprised between the “50 aa” and “100 aa” positions.

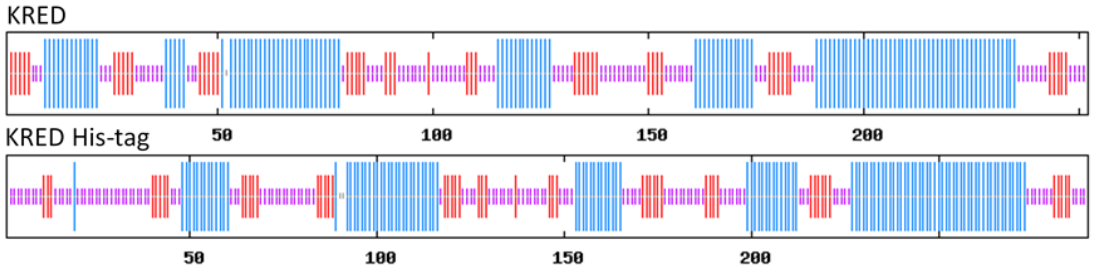


***Figure 27** Secondary structure consensus prediction of KRED and His-tag KRED.*

An online suite of predictors (SCRATCH) developed by the University of California, Irvine, was used to gather further information about the solubility of the three proteins. Two tools of the SCRATCH library were used, SOLpro and ACCpro.

SOLpro’s foundation consists on a large, non-redundant and balanced dataset of proteins expressed in *E.coli*. This dataset was prepared from the PDB, SwissProt and TargetDB databases. The final output of SOLpro consists on a prediction of solubility upon overexpression The two

possible output characterizing strings are: “SOLUBLE” or “INSOLUBLE”. Immediately after this string, another numerical string shows up revealing the probability of the first category, as in example, “SOLUBLE with probability 0.791434”. This predictor has an overall accuracy of over 74%.

SOLpro results of KRED, His-tag KRED and KRED1-pglu are displayed in Table 21. These results are in accordance with the expression studies previously stated. They also contradict the fact that KRED1-pglu has a higher hydrophobicity % than KRED which could contribute to a lower solubility.

Table 21. Solubility upon overexpression of KRED, His-tag KRED and KRED1-Pglu. Results obtained with SOLpro online tool.

Protein	Solubility upon overexpression - SOLpro
KRED	INSOLUBLE with a probability of 0.785414
His-tag KRED	INSOLUBLE with probability 0.767773
KRED1-Pglu	SOLUBLE with probability 0.756747

The ACCpro tool gives an output of the protein solvent accessibility. Solvent accessibility is the degree to which the protein residues interact with solvent molecules. The output delivers a category for each residue, "exposed"(e) or "buried"(-). This categorization is presented for different threshold values of solvent accessibility 0% (all residues exposed)-100% (all residues buried). The three proteins were evaluated at threshold values of 5%, 25% and 50%. At such threshold values each residue would be categorized as "buried" or "exposed" according to its protein’s inherent exposed surface area. ACCpro has an accuracy of up to 71% when single sequence proteins are used. The percentage of exposed residues for each protein was calculated and is presented in Table 22. No results support the His-tag insolubility in this analysis since His-tag KRED has the highest percentage of exposed residues on all of the three threshold values.

Table 22. Percentage of exposed residues of KRED, KRED His-tag and KRED1-Pglu with different solvent accessibility threshold values.

-	Percentage of exposed residues		
Protein / Solvent accessibility threshold	**5%**	**25%**	**50%**
KRED	65.48%	40.48%	10.32%
His-tag KRED	67.59%	42.07%	10.69%
KRED1-Pglu	61.90%	38.49%	10.32%

Summarizing, the His-tag fusion protein might have caused conformational changes in the KRED protein. Nevertheless, *in silico* analysis indicates that it did not increase the overall hydrophobicity of the protein or decrease the protein solvent accessibility. KRED1-Pglu has a higher hydrophobicity index and the lowest solvent accessibility. However, *in silico* analysis of solubility upon overexpression has shown that this protein is soluble and experimental evidence from the group has demonstrated that it is indeed soluble when overexpressed with a His-tag fusion protein. This suggests that other factors besides the overall hydrophobicity or the solvent accessibility may affect folding and the solubility of the His-tag KRED protein.

Chapter 4

Conclusions, limitations and future work

In this scientific project, a ketoreductase gene (*kred*) from the *Virgibacillus pantothenticus* genome was successfully cloned in frame with two different N-terminal affinity tags, namely, a six histidine fusion protein and a glutathione s-transferase fusion protein. Different expression conditions were tested in order to obtain the two recombinant proteins His-tag KRED and GST-tag KRED. The furthest point that was reached regarding the expression study of KRED was the overexpression of the His-tag KRED in an insoluble state (inclusion bodies). In a persistent effort to obtain a soluble recombinant product, which enable the determination of its enzymatic activity, a solubilisation protocol was followed. The soluble fraction obtained revealed a band with the approximate molecular weight of the His-tag KRED, suggesting that the protein might have been refolded and solubilized.

Due to the time constraints of this Erasmus+ programme, a 6 month-period was not enough to perform the cloning, expression purification and assessment of enzymatic activity of the target protein. If more time was available, more expression conditions could have been tested to reach the goal of obtaining a soluble KRED. More expression temperatures, other induction procedures and culture media could have been tested. A chaperone plasmid set could have also been tested in an attempt to improve the recombinant protein folding. Once KRED successfully expressed, with the proper folding, the last step would be to characterize its enzymatic activity.

As future work, it is expected that the previously mentioned soluble fraction is subjected to an affinity purification assay in order to verify if the His-tag KRED is present and can be successfully purified and assayed for enzymatic activity.

REFERENCES

1. **Buchholz, K., Kasche, V. and Bornscheuer, U. T.** *Biocatalysts & Enzyme Technology.* s.l. : Wiley-VCH Verlag GmbH & Co. , 2005.
2. **Faber, Kurt.** *Biotransformation in organic chemistry - A textbook, 6th edition.* s.l. : Springer, 2011.
3. **Bommarius, S. and Riebel, B. R.** *Biocatalysis.* s.l. : Wiley-VCH Verlag GmbH & Co, 2004.
4. **Donadio, S. and Marino, G.** *Biotecnologie microbiche.* s.l. : CEA, 2008.
5. **Engineering, Committee on Bioprocess, et al.** *Putting Biotechnology to Work: Bioprocess Engineering.* s.l. : National Research Council, 1992.
6. *Protein refolding for industrial processes.* **Clark, Eliana De Bernardez.** s.l. : Current Opinion in Biotechnology, 2001, Vol. 12.
7. *Gene Expression Systems: Using nature for the art of expression.* **Fernandez, J.M. and Hoeffler, J.P.** s.l. : Academic Press, 1999.
8. *Posttranslational modification of therapeutic proteins in plants.* **Gomord, Veronique and Faye, Loïc.** s.l. : Current Opinion in Plant Biology, 2004, Vol. 7.
9. *(Microbial ethanol production: Experimental study and multivariate evaluation .* **A., Boumba V., et al.** s.l. : Forensic Science International, 2012.
10. *Recombinant protein expression in Escherichia coli.* **F., Baneyx.** s.l. : Current Opinion in Biotechnology , 1999, Vol. 10.
11. *Systematic high-yield production of human secreted proteins in Escherichia coli.* **Xueyu, Daia, et al.** s.l. : Biochemical and Biophysical Research Communications, 2005, Vol. 332.
12. *Marine metagenomics: strategies for the discovery of novel enzymes with biotechnological applications from marine environments.* **Kennedy, J., Marchesi, J. R. and DW., D. A.** s.l. : Microbial Cell Factories, 2008.
13. *Novel molecular methods for discovery and engineering of biocatalysts from uncultured marine microorganisms.* **Uria, A. R., Nuri, F. Y. and Ekowati, C.** s.l. : Journal of Coastal Development, 2005, Vol. 8.
14. *Marine enzymes.* **Debashish, G., et al.** s.l. : Advances in Biochemical Engineering/Biotechnology, 2005, Vol. 96.

15. *Enzymes from extremophiles.* **Demirjian, D.C., Morís-Varas, F. and Cassidy, C.S.** s.l. : Current opinion in chemical biology, 2001, Vol. 5.
16. *Marine biominerals: perspectives and challenges for polymetallic.* **Wang, X. and Müller, W.E.G.** s.l. : Trends in Biotechnology, 2009, Vol. 27.
17. *Fish bioaccumulation and biomarkers in environmental risk assessment: a review.* **van der Oost, R., Beyer, J. and Vermeulen, N.P.E.** s.l. : Environmental Toxicology and Pharmacology, 2003, Vol. 13.
18. *Identification of proteomic signatures of exposure to marine pollutants in mussels (Mytilus edulis).* **Apraiz, I. and Cristobal, S.** s.l. : Molecular and Cellular Proteomics, 2006, Vol. 5.
19. *Pseudoalteromonas bacteria are capable of degrading paralytic shellfish toxins.* **Donovan, C.J., et al.** s.l. : Applied and Environmental Microbiology, 2009, Vol. 75.
20. *Polyketide synthases of bacterial symbionts in sponges—Evolution-based applications in natural products research.* **Hochmuth, T. and Piel, J.** s.l. : Phytochemistry, 2009, Vol. 70.
21. *Approaches for novel enzyme discovery from marine environments.* **Lee, H.S., et al.** s.l. : Current Opinion in Biotechnology, 2010, Vol. 21.
22. *Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism.* **Kennedy, J., et al.** s.l. : Marine Drugs, 2010, Vol. 8.
23. *Finding new enzymes from bacterial physiology: a successful approach illustrated by the detection of novel oxidases in Marinomonas mediterranea.* **Sanchez-Amat, A., Solano, F. and Lucas-Elío, P.** s.l. : Marine drugs, 2010, Vol. 8.
24. *Extreme halophilic enzymes in organic solvents.* **Marhuenda-Egea, F.C. and Bonete, M.J.** s.l. : Current Opinion in Biotechnology, 2002, Vol. 13.
25. *Biochemical properties of anionic trypsin acting at high concentration of NaCl purified from the intestine of a carnivorous fish: smooth hound (Mustelus mustelus).* **Bougatef, A., et al.** s.l. : Journal of Agricultural and Food Chemistry, 2010, Vol. 58.
26. *Saccharification of marine microalgae using marine bacteria for ethanol production.* **Matsumoto, M., et al.** s.l. : Applied Biochemistry and Biotechnology, 2003.
27. *Industrial relevance of thermophilic Archaea.* **Egorova, K. and Antranikian, G.** s.l. : Current Opinion in Microbiology, 2005, Vol. 8.
28. *Asymmetric ketone reduction by a hyperthermophilic alcohol dehydrogenase. The substrate specificity, enantioselectivity and tolerance of organic solvents.* **Zhu, D., Malik, H.T. and Hua, L.** s.l. : Tetrahedron: Asymmetry, 2006, Vol. 17.
29. *Characterization of a thermostable β-glucosidase (BglB) from Thermotoga maritima showing transglycosylation activity.* **Goyal, K., Selvakumar, P. and Hayashi, K.** s.l. : Journal of Molecular Catalysis B: Enzymatic, 2001, Vol. 15.
30. *The recombinant xylanase B of Thermotoga maritima is highly xylan specific and produces exclusively xylobiose from xylans, a unique character for industrial applications.* **Jiang, Z.Q., et al.** s.l. : Journal of Molecular Catalysis B: Enzymatic, 2004, Vol. 27.
31. *Transglycosylation reactions using glycosyl hydrolases from Thermotoga neapolitana, a marine hydrogen-producing bacterium.* **Tramice, A., et al.** s.l. : Journal of Molecular Catalysis B: Enzymatic, 2007, Vol. 47.
32. *Microbial life at high pressures.* **Bartlett, D.H.** s.l. : Science Progress, 1992, Vol. 76.

33. *Barophiles: deep-sea microorganisms adapted to an extreme environment.* **Horikoshi, K.** s.l. : Current Opinion in Microbiology, 1998, Vol. 1.
34. *Are the catalytic properties of enzymes from piezophilic organisms pressure adapted?* **Hay, S., et al.** s.l. : ChemBioChem, 2009, Vol. 10.
35. *Pressure and temperature effects on growth and viability of the hyperthermophilic archaeon Thermococcus Peptonophilus.* **Canganella, F., et al.** s.l. : Archives of Microbiology, 1997, Vol. 168.
36. *Biosynthesis and dietary uptake of polyunsaturated fatty acids by piezophilic bacteria.* **Fang, J., et al.** s.l. : Comparative Biochemistry and Physiology, 2004, Vol. 137.
37. *Lipolytic activity of Antarctic cold-adapted marine bacteria (Terra Nova Bay, Ross Sea).* **Lo Giudice, A., et al.** s.l. : Journal of Applied Microbiology , 2006, Vol. 101.
38. *A cold-adapted esterase of a novel marine isolate, Pseudoalteromonas arctica: gene cloning, enzyme purification and characterization.* **Khudary, R.A., et al.** s.l. : Extremophiles, 2010, Vol. 14.
39. *Biocatalytic resolution of glycidyl phenyl ether using a novel epoxide hydrolase from a marine bacterium, Rhodobacterales bacterium HTCC2654.* **Woo, J.-H., et al.** s.l. : Journal of Bioscience and Bioengineering, 2010, Vol. 109.
40. *Enantioselective epoxide hydrolase activity of a newly isolated microorganism, Sphingomonas echinoides EH-983, from seawater.* **Kim, H.S., et al.** s.l. : Journal of Molecular Catalysis B: Enzymatic, 2006, Vol. 41.
41. *Biotransformation of polycyclic aromatic hydrocarbons in marine polychaetes.* **Jorgensen, A., et al.** s.l. : Marine Environmental Research , 2008, Vol. 65.
42. *Hydroxylations of substituted naphthalenes by Escherichia coli expressing aromatic dihydroxylating dioxygenase genes from polycyclic aromatic hydrocarbon-utilizing marine bacteria.* **Shindo, K., et al.** s.l. : Journal of Molecular Catalysis B: Enzymatic, 2007, Vol. 48.
43. *Marine Biocatalysts: Enzymatic Features and Applications.* **Trincone, Antonio.** s.l. : Marine Drugs, 2011.
44. *Production of a recombinant alkane hydroxylase (AlkB2) from Alcanivorax borkumensis.* **Miri, M., et al.** s.l. : Biotechnology Letters, 2010, Vol. 32.
45. *A novel β-Agarase with high pH stability from marine Agarivorans sp. LQ48.* **Long, M., Yu, Z. and Xu, X.** s.l. : Marine Biotechnology, 2010, Vol. 12.
46. *A unique β-1,2-mannosyltransferase of Thermotoga maritima that uses di-myo-inositol phosphate as the mannosyl acceptor.* **Rodrigues, M.V., et al.** s.l. : Journal of Bacteriology, 2009, Vol. 191.
47. *Screening and its potential application of lipolytic activity from a marine environment: characterization of a novel esterase from Yarrowia lipolytica CL180.* **Kim, J.-T, et al.** s.l. : Applied Microbiology and Biotechnology, 2007, Vol. 74.
48. *Application of linoleic acid hydroperoxide as a mild and green bleaching agent.* **Cai, Y., et al.** s.l. : AIChE Annual Meeting.
49. *Enantioselective formation of (R)-9-HPODE and (R)-9-HPOTrE in marine green alga Ulva conglobata.* **Akakabe, Y., Matsui, K. and Kajiwara, T.** s.l. : Bioorganic & Medicinal Chemistry Letters, 2002, Vol. 10.
50. *Convenient synthesis of β-galactosyl nucleosides using the marine β-galactosidase from Aplysia fasciata.* **Andreotti, G., Trincone, A. and Giordano, A.** s.l. : Journal of Molecular Catalysis - B Enzymatic, 2007, Vol. 47.

51. *Evidence for isolated evolution of deep-sea ciliate communities through geological separation and environmental selection.* **Edgcomb, V., et al.** s.l. : BMC MICROBIOLOGY ·, 2013.
52. *Anoxic Basins of the Eastern Mediterranean: an overview.* **Cita, M. B.** s.l. : PALEOCEANOGRAPHY, 1991, Vol. 6.
53. *A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding.* **Bradford, M. M.** s.l. : ANALYTICAL BIOCHEMISTRY , 1976, Vol. 72.
54. EMBL - European Molecular Biology Laboratory. [Online] https://www.embl.de/pepcore/pepcore_services/protein_purification/extraction_clarification/solubilisation_renaturation/.
55. *Effective renaturation of denatured and reduced immunoglobulin G in vitro without assistance of chaperone.* **Maeda, Y., Ueda, T. and Imoto, T.** s.l. : Protein Engineering, 1996, Vol. 9.
56. Tirabassi, Dr Rebecca. BitsizeBio. [Online] http://bitesizebio.com/13524/how-to-identify-supercoils-nicks-and-circles-in-plasmid-preps/.
57. *Low temperatures stabilize interferon α-2 against proteolysis in Methylophilus methylotrophus and Escherichia coli.* **Chesshyre, Julian A. and Hipkiss, Alan R.** s.l. : Applied Microbiology and Biotechnology.
58. *Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation.* **Kiefhaber, T., et al.** s.l. : Biotechnology (Nature Publishing Company), 1991.
59. *Mechanisms of protein folding: molecular chaperones and their application in biotechnology.* **Mogk, Axel., Mayer, Matthias. P. and Deuerling, Elke.** s.l. : ChemBioChem, 2002.
60. *Chaperonins govern growth of Escherichia coli at low temperatures.* **Ferrer, Manuel, et al.** s.l. : Nature Biotechnology, 2003.
61. *Synthesis of macromolecules by Escherichia coli near the minimal temperature for growth.* **Shaw, Maxwell K. and Ingraham, John L.** s.l. : Journal of Bacteriology, 1967.
62. *Recombinant protein expression at low temperatures under the transcriptional control of the major Escherichia coli cold shock promoter cspA.* **Vasina, Jess A. and Baneyx, François.** s.l. : Applied and Environmental Microbiology, 1996.
63. *Optimization of heterologous protein production in Escherichia coli.* **Weickert, Michael J., et al.** s.l. : Current Opinion in Biotechnology, 1996.
64. *Purification of proteins fused to glutathione S-tranferase.* **Harper, Sandra and Speicher, David W.** s.l. : Methods in Molecular Biology, 2011.
65. *Modifications of therapeutics proteins: challenges and prospects.* **Jenkins, Nigel.** s.l. : Cytotechnology, 2007.
66. *Post-translational modifictions in the context of therapeutic proteins.* **Walsh, Gary and Jefferis, Roy.** s.l. : Nature Biotechnology, 2006.
67. *Stereoselective reduction of aromatic ketones by a new ketoreductase from Pichia glucozyma.* **Contente, Martina Letizia, et al.** s.l. : Applied Microbiology and Biotechnology, 2016, Vol. 100.
68. SAPS - Statistical Analysis of Protein Sequences . [Online] http://www.ebi.ac.uk/Tools/seqstats/saps/.

69. Prabi-Gerland Rhone-Alpes Bioinformatic Pole Gerland Site. [Online] https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html.

▪ ANNEX 1

***Table 23**. Reagents (Chemistry and Molecular Biology) and solutions used throughout this project*

REAGENTS				
NAME	IDENTIFICATION NUMBER	BRAND	MOLECULAR WEIGHT (g/mol)	DESCRIPTION
2-Mercaptoethanol (C_2H_6OS)	CAS: 60-24-2	BDH	78.13	
3-(N-Morpholino)propanesulfonic acid ($C_7H_{15}NO_4S$)	CAS: 1132-61-2	Sigma - Aldrich	209.26	
Acetic Acid (CH_3CO_2H)	CAS: 64-19-7	Sigma - Aldrich	60.05	
Ammonium Persulfate ($H_8N_2O_8S_2$)	CAS: 7727-74-0	BDH	228.2	
Ammonium sulfate$(NH_4)_2SO_4$	CAS: 7783-20-2	Sigma - Aldrich	132.14	
Brilliant Blue R ($C_{45}H_{44}N_3NaO_7S_2$)	CAS: 6104-59-2	Sigma - Aldrich	825.97	
Bromophenol Blue	CAS: 115-39-9	Sigma - Aldrich	669.96	
Calcium chloride ($CaCl_2$)	CAS: 10043-52-4	Sigma - Aldrich	110.98	
Chloramphenicol ($Cl_2CHCONHCH(CH_2OH)CH(OH)C_6H_4NO_2$)	CAS: 56-75-7	Sigma - Aldrich	323.13	
Ethylenediaminetetraacetic acid (EDTA)	CAS: 60-00-4	Sigma - Aldrich	292.24	

Ethanol (C_2H_5OH)	CAS: 64-17-5	Sigma - Aldrich	46.07	
Ethidium bromide ($C_{21}H_{20}BrN_3$)	CAS: 1239-45-8	Sigma - Aldrich	394.31	
Glycerol ($HOCH_2CH(OH)CH_2OH$)	CAS: 56-81-5	SAFC	92.09	
Glycine (NH_2CH_2COOH)	CAS: 56-40-6	Sigma - Aldrich	75.07	
Imidazole	CAS: 288-32-4	Sigma - Aldrich	68.08	
Isopropyl ß-D-1-thiogalactopyranoside (IPTG)	CAS: 367-93-1	Sigma - Aldrich	238.30	
Magnesium SulfateHeptahydrate ($MgSO_4 \cdot 7H_2O$)	CAS: 1034-99-8	Sigma - Aldrich	246.47	
Monobasic potassium phosphate (KH_2PO_4)	CAS: 7778-77-0	Sigma - Aldrich	136,09	
N,*N*,*N*′,*N*′-Tetramethylethylenediamine (TEMED) ($(CH_3)_2NCH_2CH_2N(CH_3)_2$)	CAS: 110-18-9	Sigma - Aldrich	116.20	
Rubidium Chloride	CAS: 7791-11-9	Sigma - Aldrich	120.92	
Sodium Chloride	CAS: 7647-14-5	Sigma - Aldrich	58.44	
Sodium dodecyl sulfate (SDS, $(CH_3(CH_2)_{11}OSO_3Na)$	CAS: 151-21-2	Sigma - Aldrich	288.38	
Sodium phosphate dibasic (Na_2HPO_4)	CAS: 7558-79-4	Sigma - Aldrich	141,96	
Sodium phosphate monobasic (NaH_2PO_4)	CAS: 7558-80-7	Sigma - Aldrich	119,98	
Sodium Sulfate (Na_2SO_4)	CAS: 7757-82-6	Sigma - Aldrich	142.04	

Streptomycin sulfate salt ($C_{21}H_{39}N_7O_{12} \cdot 1.5\ H_2SO_4$)	CAS: 3810-74-0	Sigma - Aldrich	728-69	
Tris(hydroxymethyl)aminomethane (Tris)	CAS: 77-86-1	Sigma - Aldrich	121.14	
Tryptone	CAS: 91079-40-2	BiolifeItaliana		
Yeast Extract		BiolifeItaliana		Yeast Extract is a water-soluble spry-dried extract resulting from the autolysis of primary grown yeast cells of the species Saccharomyces cerevisiae. During the production the process is carefully controlled so that the B-complex vitamins are almost completely conserved. Yeast Extract is an excellent source of B-complex vitamins and it is standardised for bacteriological purposes.

α-Lactose Monohydrate	CAS: 5989-81-1	Sigma - Aldrich	360.31	
MOLECULAR BIOLOGY REAGENTS				
NAME	IDENTIFICATION NUMBER	BRAND	DESCRIPTION	
Acrylamide/Bis-acrylamide, 30% solution		Sigma - Aldrich	The Acrylamide/Bis-acrylamide solution is used in protein and nucleic acid electrophoresis. The solution concentration (30%) is based on the total weight of both the acrylamide and bis-acrylamide.	
D-1 Agarose	CAT: 8010.00	CONDA Pronadisa	Low Electroendosmosis (EEO): 0,5 – 0,13 High electrophoresis mobility; Nucleic acid analytical and preparative electrophoresis.; Blotting; Protein electrophoresis such as radial immunodiffusion	
DNA Gel Loading Die (6x)	Catalog number: R0611	Thermo Fisher Scientific Inc.		
DreamTaq DNA Polymerase (5 U/µL)	Catalog number:EP0701	Thermo Fisher Scientific Inc.		
DreamTaq Green Buffer (10X)	Catalog number: B71	Thermo Fisher Scientific Inc.		

FastAP Thermosensitive Alkaline Phosphatase (1 U/µL)	Catalog number: EF0654	Thermo Fisher Scientific Inc.	Thermo Scientific FastAP Thermosensitive Alkaline Phosphatase catalyzes the release of 5'- and 3'-phosphate groups from DNA, RNA, and nucleotides. This enzyme also removes phosphate groups from proteins.
FastDigestBamHI	Catalog number: FD0054	Thermo Fisher Scientific Inc.	Thermo Scientific FastDigestBamHI restriction enzyme recognizes G^GATCC site and cuts best at 37°C in 5–15 minutes using universal FastDigest Buffer.
FastDigest Buffer (10X)	Catalog number: B64	Thermo Fisher Scientific Inc.	The 10X FastDigest Buffer is a proprietary digestion buffer which supports 100% activity of all FastDigest restriction enzymes.
FastDigestEcoRI	Catalog number: FD0274	Thermo Fisher Scientific Inc.	Thermo Scientific FastDigestEcoRI restriction enzyme recognizes G^AATTC site and cuts best at 37°C in 5–15 minutes using universal FastDigest Buffer.
FastDigest Green Buffer (10X)	Catalog number: B72	Thermo Fisher Scientific Inc.	The 10X FastDigest Green Buffer is a proprietary digestion buffer which supports 100% activity of all FastDigest restriction enzymes.

FastDigestHindIII	Catalog number: FD0504	Thermo Fisher Scientific Inc.	Thermo Scientific FastDigestHindIII restriction enzyme recognizes A^AGCTT site and cuts best at 37°C in 5–15 minutes using universal FastDigest Buffer.
FastDigestXhoI	Catalog number: FD0694	Thermo Fisher Scientific Inc.	Thermo Scientific FastDigestXhoI restriction enzyme recognizes C^TCGAG site and cuts best at 37°C in 5–15 minutes using universal FastDigest Buffer.
Gene Ruler 1 Kb	Catalog number: SM0312	Thermo Fisher Scientific Inc.	
HIS-Select® Nickel Affinity Gel	Catalog number: P6611	Sigma - Aldrich	Nickel Affinity Gel used for the purification of histidine-tagged proteins. The unique, non-charged, hydrophilic linkage of the NTA chelate group to the agarose ensures high selectivity for histidine-tagged proteins and reduced non-specific binding of other proteins. Its high selectivity allows for one-step purification. Affinity gel can be used in either denaturing or non-denaturing conditions
Phusion High-Fidelity DNA Polymerase (2 U/µL)	Catalog number: F530S	Thermo Fisher	

		Scientific Inc.	
Pierce™ Prestained Protein MW Marker	Catalog number: 26612	Thermo Fisher Scientific Inc.	
SigmaMarker S8445	Catalog number: S8445	Sigma - Aldrich	• Wide range: 6,500-200,000 Da
T4 DNA Ligase (5 U/µL)	Catalog number: EL0014	Thermo Fisher Scientific Inc.	Thermo Scientific T4 DNA Ligase catalyzes the formation of a phosphodiester bond between juxtaposed 5'-phosphate and 3'-hydroxyl termini in duplex DNA or RNA. The enzyme repairs single-strand nicks in duplex DNA, RNA, or DNA/RNA hybrids. It also joins DNA fragments with either cohesive or blunt termini, but has no activity on single-stranded nucleic acids.

SOLUTIONS AND CULTURE MEDIA

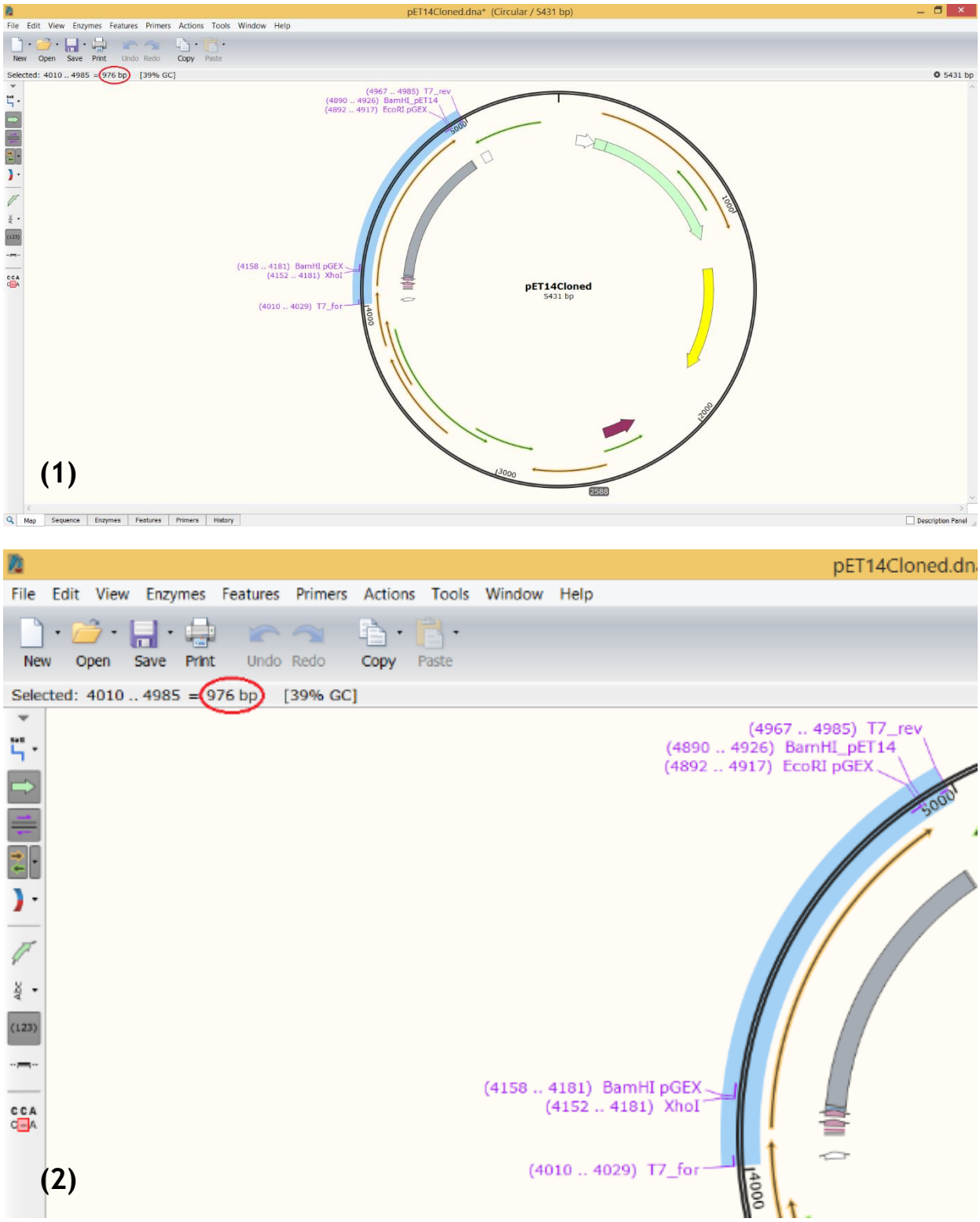
NAME	RECIPE
Auto-Induction medium (ZYM-5052)	Na_2HPO_4 25 mM KH_2PO_4 25 mM NH_4Cl 50 mM Na_2SO_4 5 mM Glycerol 0.5% Gucose 0.05% α-lactose 0.2% Tryptone 1%

	Yeast extract 0.5%
Bromophenol Blue solution	20 mg Bromophenol Blue 6 mg Tris 40 mL dH_2O
Comassie Blue Solution	For 100 mL: • 0,25 g Brilliant Blue R • 50 mL Ethanol • 40 mL dH2O • 10 mL Acetic Acid
DEN+	For 8 mL: • 300 mg SDS • 2 mL dH_2O • 2 mL Blue bromophenol • 200 µL 2-Mercaptoethanol • 3,8 mL Glycine
DEN + Concentrated	For 4 mL: • 0,1 mL Resolving Buffer pH 8.8 • 0,5 mL Stacking Buffer pH 6.8 • 3 mg Blue bromophenol • 500 mg SDS • 3,4 mL Saturated sacarose solution (dH_2O+sacarose)
DNA marker	10 µL Gene Ruler 1 Kb 10 µL DNA Gel Loading Die (6x) 40 µL dH_2O
Extraction Buffer	50 mMTris-HCl pH 7.5 8 M urea 1 mM DTT (Dithiothreitol)
Imidazole Buffer 6mM pH 8.0	50 mMTris-HCl pH 8.0 100 mMNaCl 6 mM Imidazole
Imidazole Buffer 250mM pH 8.0	50 mMTris-HCl pH 8.0 100 mMNaCl 250 mM Imidazole
Lysogeny Broth (LB)	10 g/L NaCL 5 g/L Yeast Extract 10 g/L Tryptone dH_2O
Page Running Buffer 10X	30 g Tris 140 g Glycine 5% SDS

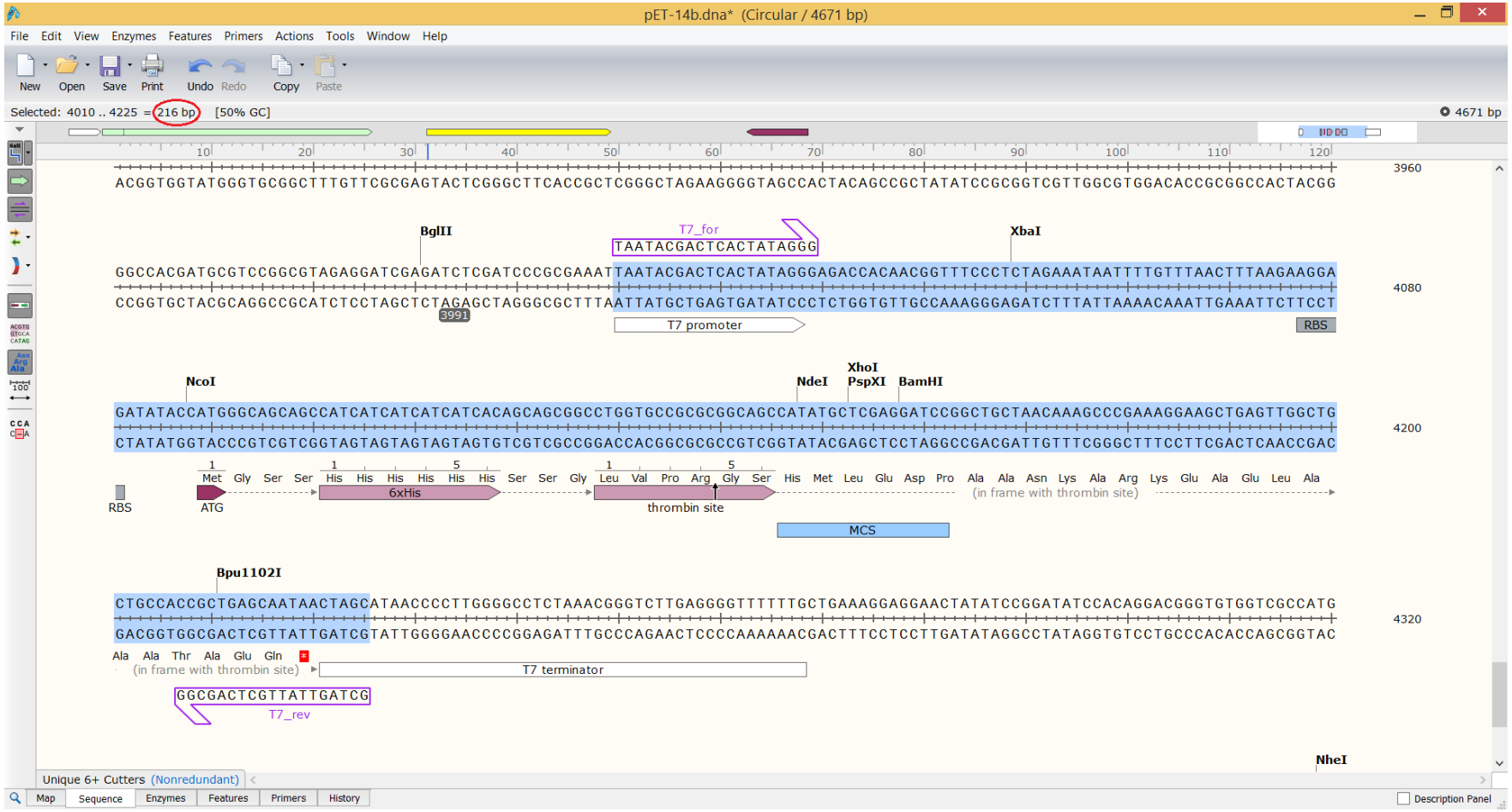
Phosphate Buffer Saline (PBS) pH 7.0	Add a 100 mM NaH_2PO_4 solution to a 100 mM Na_2HPO_4 until the pH reaches 7.0
Resolving Buffer pH 8.8	1,5 M Tris-HCl pH 8.8 0,4% SDS
Stacking Buffer pH. 6.8	0,5 M Tris-HCl pH 8.8 0,4% SDS
Tris-Acetate EDTA buffer (TAE) 50X	242 g Tris 57.1 mL acetic acid 500 mM EDTA 0,5 mM (pH 8.0) dH_2O to perform 1 L
Wash Buffer	50 mMTris-HCl pH 7.5 100 mMNaCl

NOTE: CAS - Chemical Abstracts Service

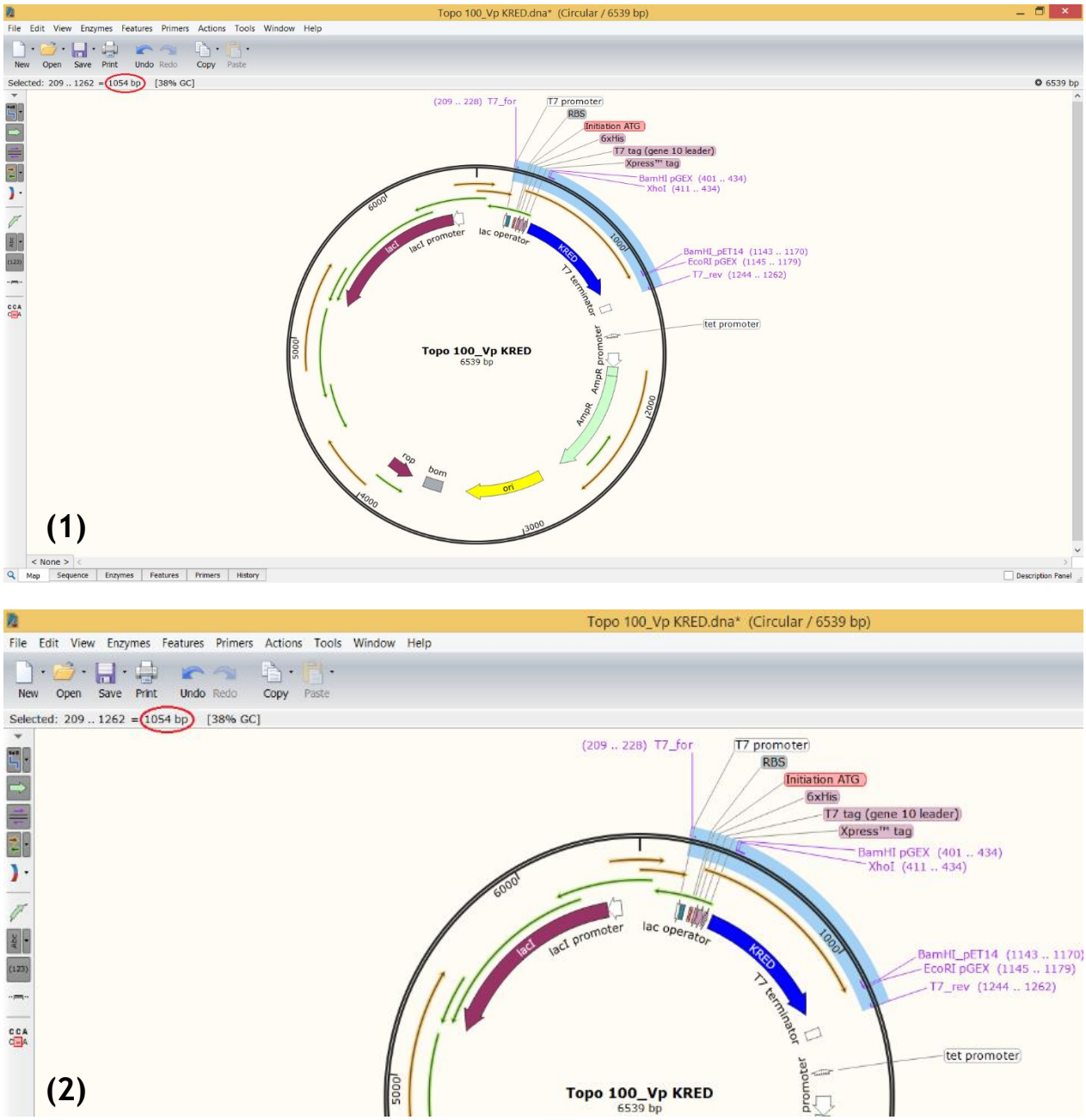
▪ ANNEX 2



*PICTURE A. **(1)** Length measurement of the encoded fragment between the T7 promoter primers in the kred pET-14b vector. **(2)** Zoom of picture (1).*

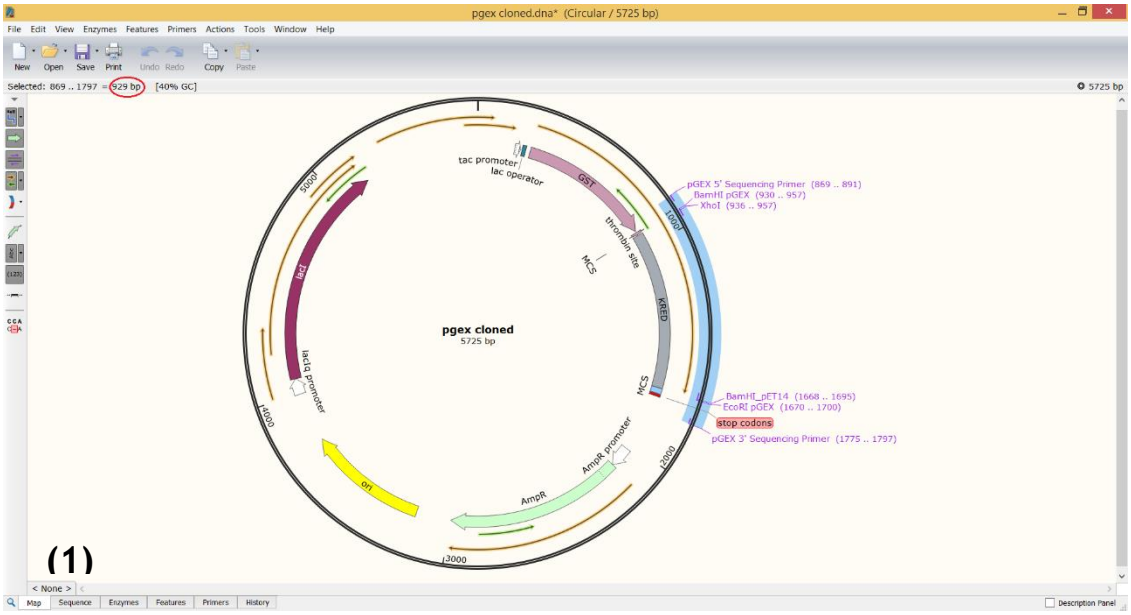


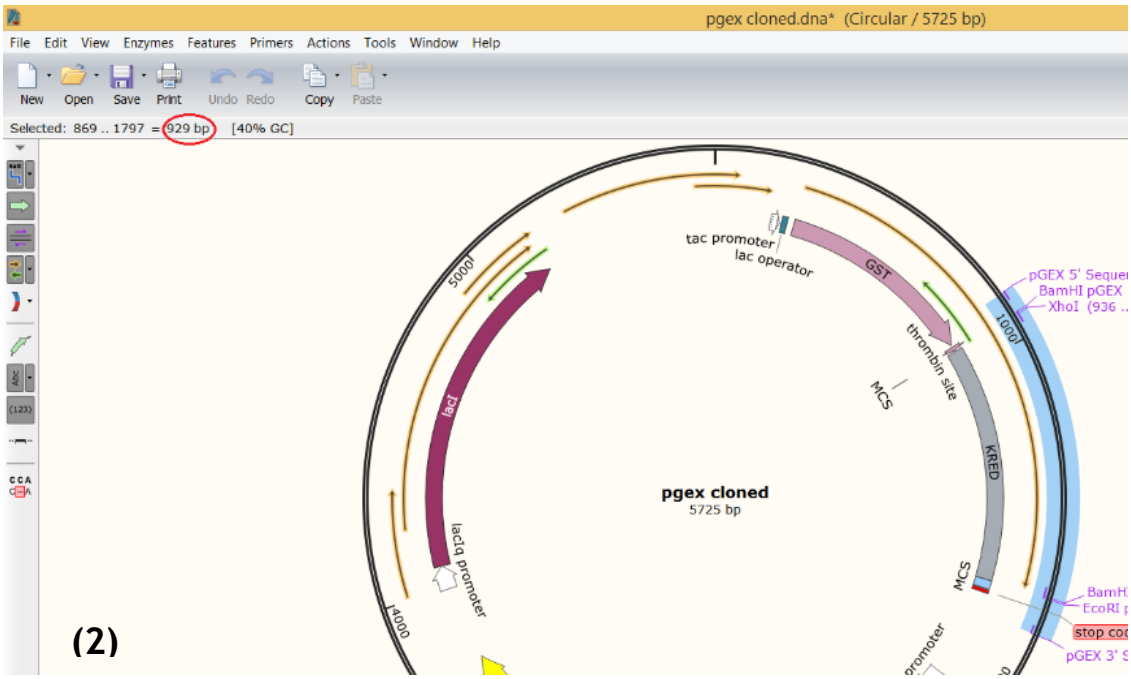
PICTURE B. *Length measurement of the encoded fragment between the T7 promoter primers in the pET-14b vector.*



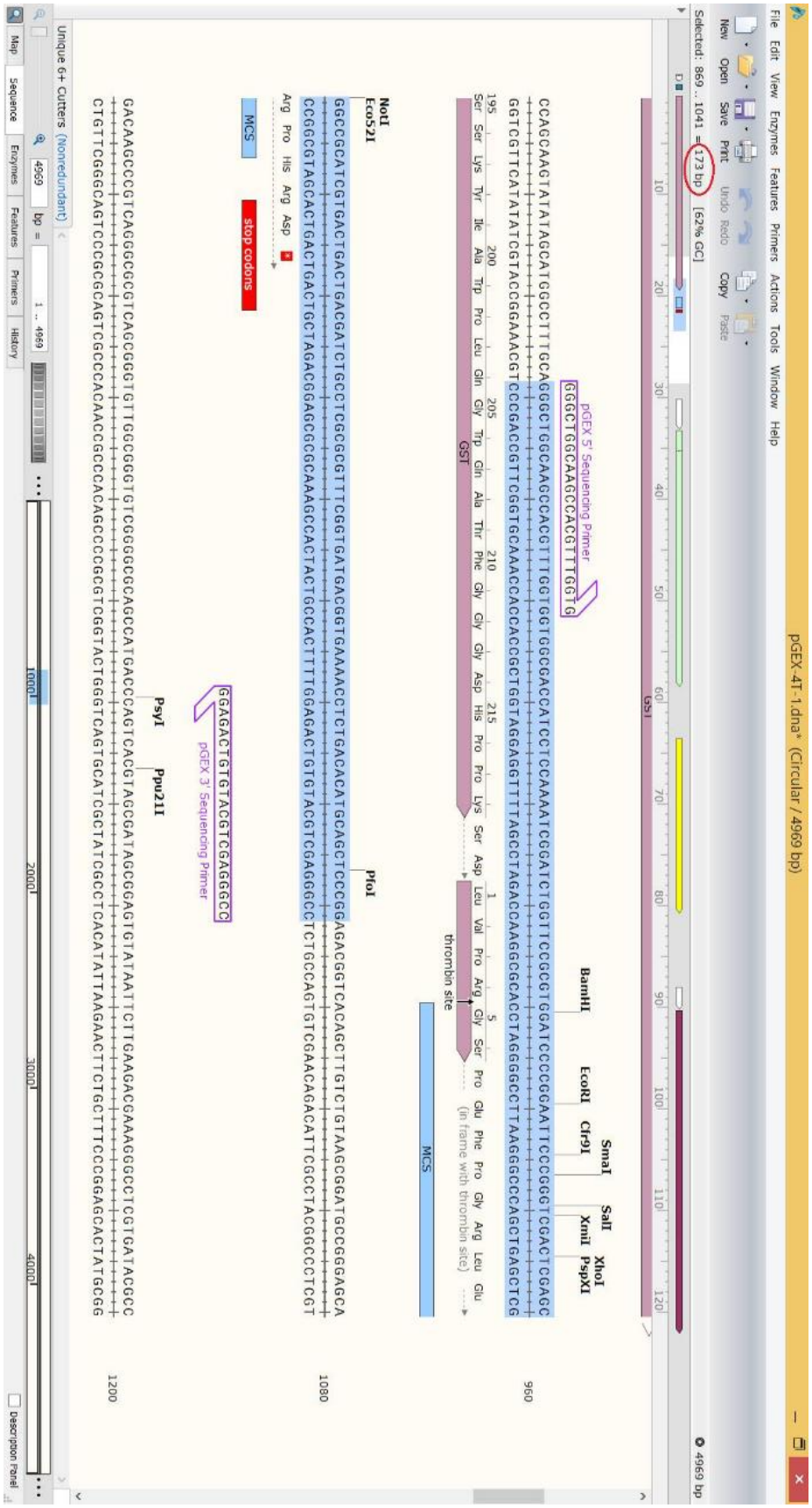
PICTURE C. *(1) Length measurement of the encoded fragment between the T7 promoter primers in the kred pET-100 TOPO vector.* ***(2)*** *Zoom of picture (1).*

***PICTURE D**. Length measurement of the encoded fragment between the T7 promoter primers in the pET-100 TOPO vector*





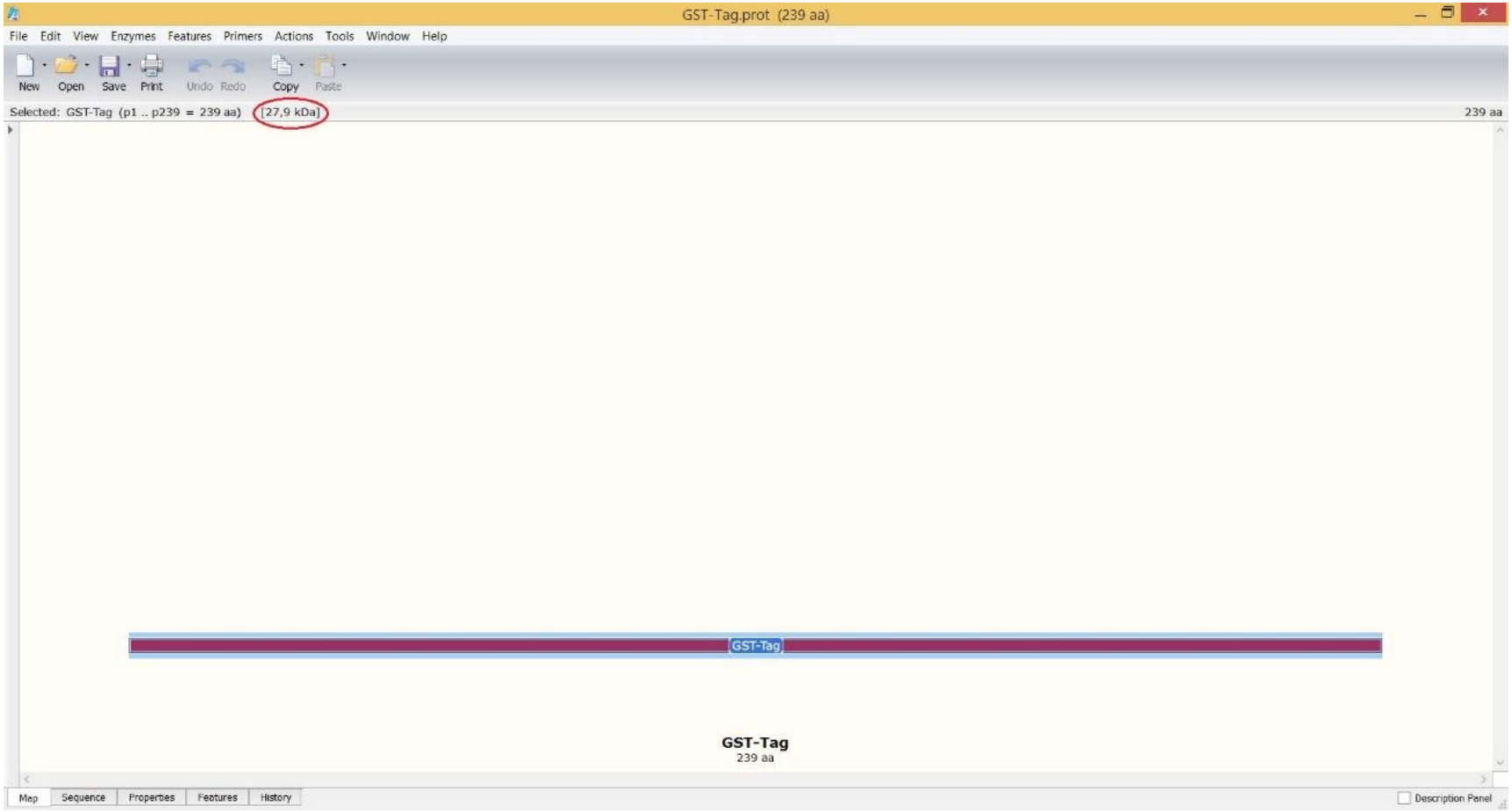
PICTURE E. *(1) Length measurement of the encoded fragment between the pGEX 5' and 3' primers in the kred pGEX-4T-1 vector.* ***(2)*** *Zoom of picture (1).*



***PICTURE F**. Length measurement of the encoded fragment between the pGEX 5' and 3' primers in the pGEX-4T-1 vector.*

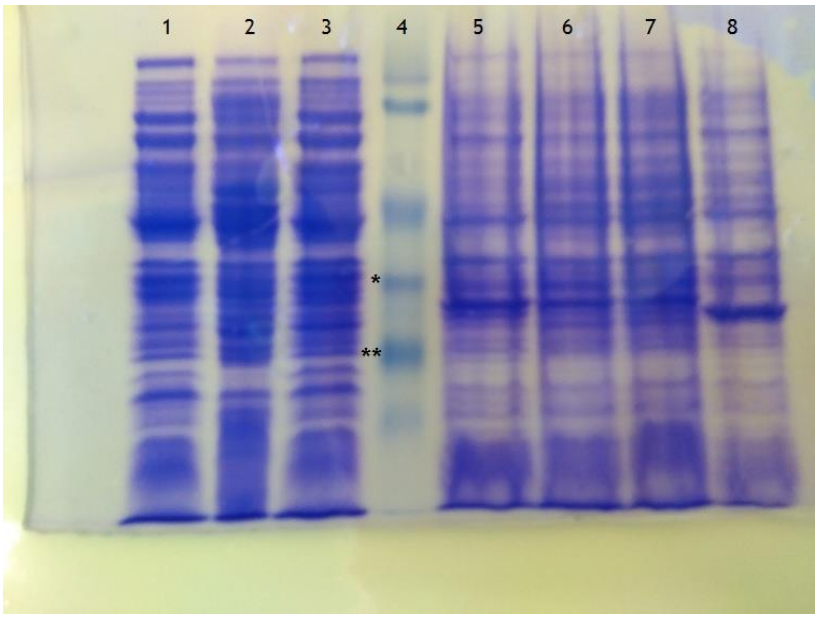
PICTURE G. His-tag KRED molecular weight measurement.

PICTURE H. GST-tag KRED molecular weight measurement.

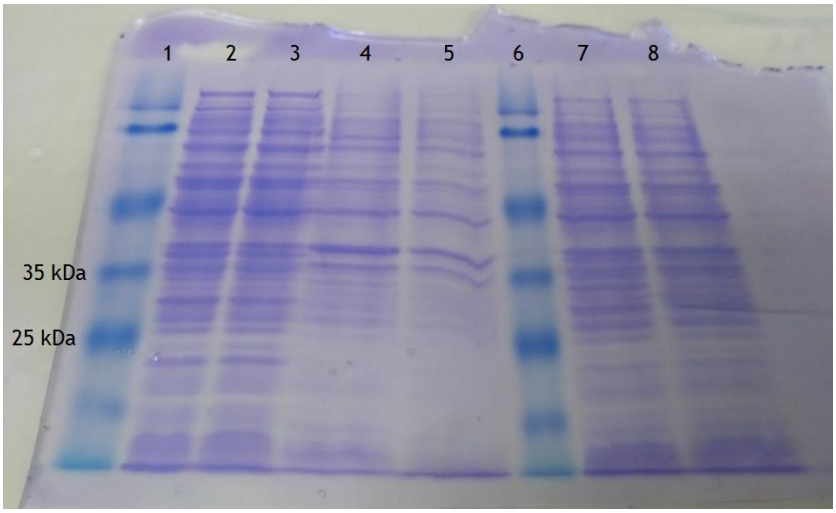


PICTURE I. GST fusion protein molecular weight measurement.

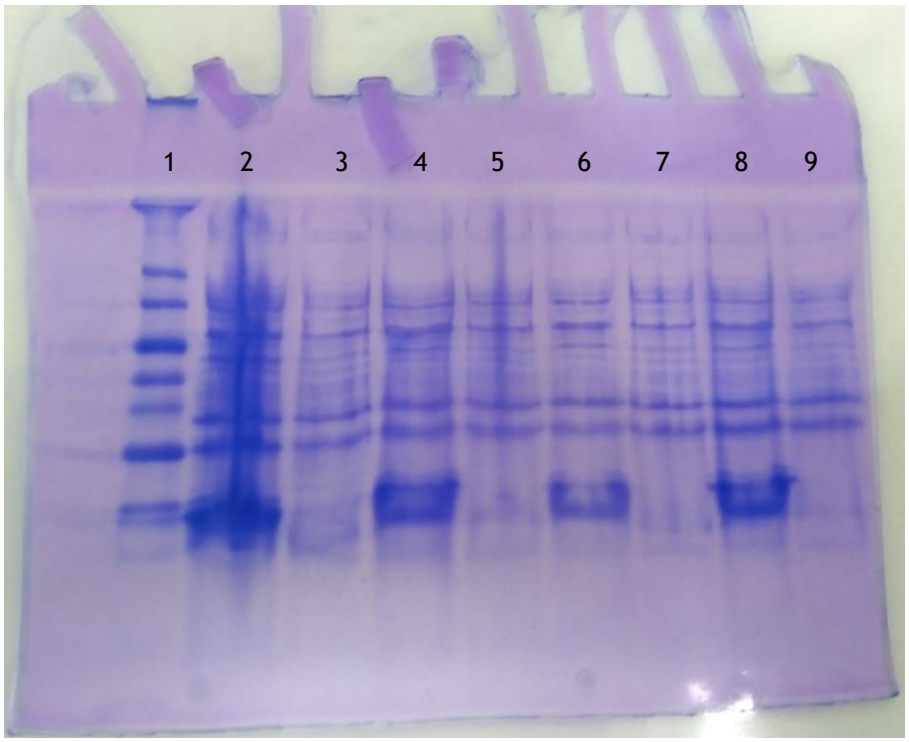
▪ ANNEX 3



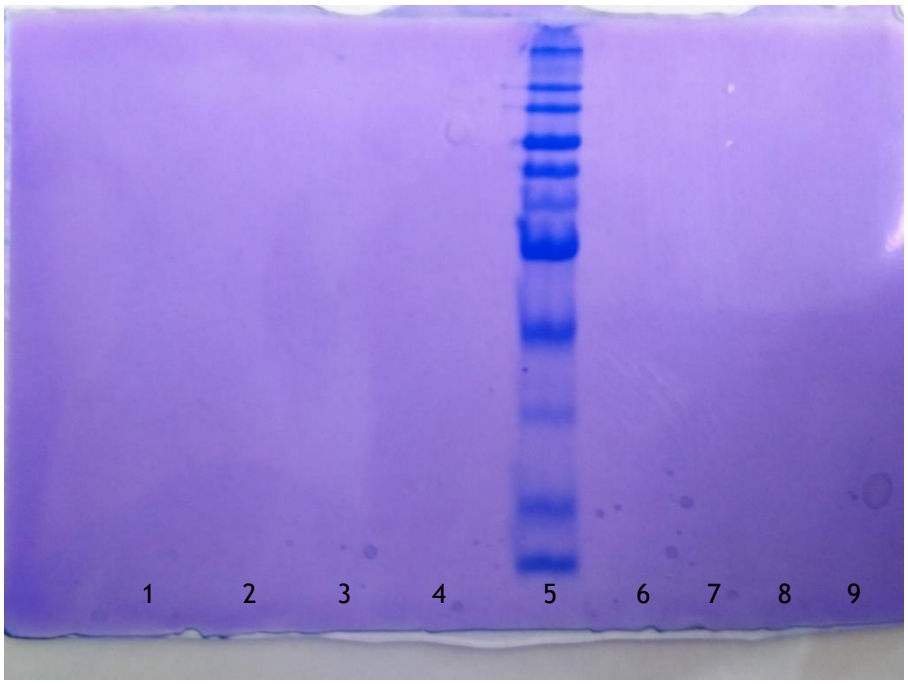
***PICTURE J**. SDS-PAGE of expression study of kred pET-100 TOPO with BL21(DE3) STAR™ at 25ºC with cold-shock. Lane 1 - Cell extract kred pET-100 TOPO, 3h IPTG induction ; Lane 2 - Cell extract kred pET-100 TOPO, 20h IPTG induction ; Lane 3 - Cell extract pET-100 TOPO, 20h IPTG induction ; Lane 4 - Pierce™ Prestained Protein MW Marker, * - 35 kDa, ** - 25 kDa ; Lane 5 - Whole cells kred pET-100 TOPO, 3h IPTG induction ; Lane 6 - Whole cells kred pET-100 TOPO, 20h IPTG induction; Lane 7 - Whole cells kred pET-100 TOPO, non-induced; Lane 8 - Whole cells pET-100 TOPO, non-induced.*



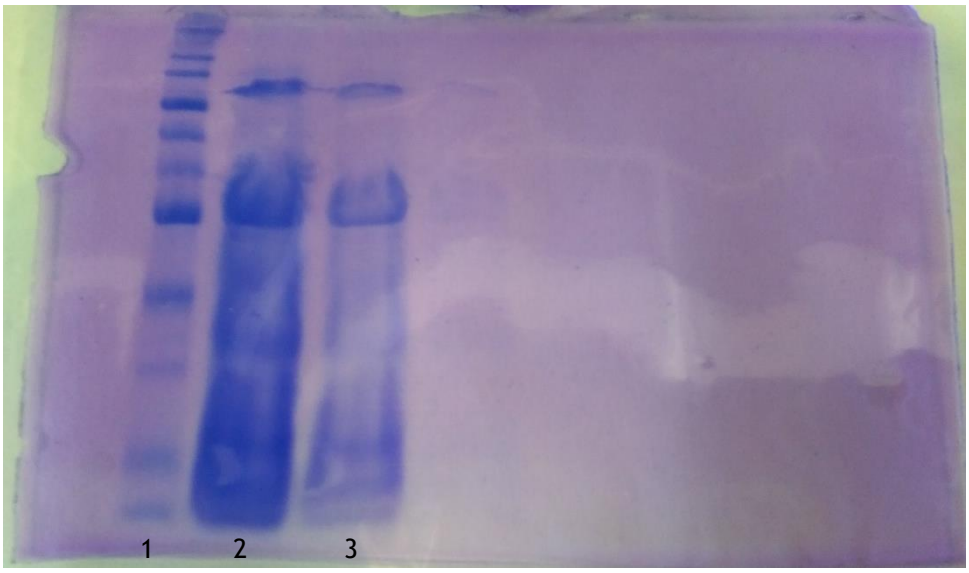
***PICTURE K**.* SDS-PAGE of expression study of kred pET-100 TOPO with BL21(DE3) STAR™ at 37 ºC in auto-induction (AI) medium. Lanes 1 and 6 - Pierce™ Prestained Protein MW Marker; Lane 2 - Cell extract kred pET-100 TOPO, 24h; Lane 3 - Cell extract pET-100 TOPO, 24h; Lane 4 - Sonicated pellet kred pET-100 TOPO, 24h; Lane 5 - Sonicated pellet pET-100 TOPO; Lane 7 - Whole cells kred pET-100 TOPO, 24h; Lane 8 - Whole cells pET-100 TOPO, 24h.



***PICTURE L**. Screening of kred pGEX-4T-1 (CODON+) transformants. This gel presents 4 out of the 6 screened colonies. Lane 1 - SigmaMarker S8445; Lanes 2 & 3, 4 & 5, 6 & 7, 8 & 9 – the first lane of each pair corresponds to an induced culture, while the second one corresponds to a non-induced culture.*



***PICTURE M**. SDS-PAGE of the affinity purification assay of kred pET-100 TOPO 20h induction, 25ºC. The eight eluted fractions are presented in this gel. Lanes 1 to 4 and 6 to 9 - eight eluted fractions, loaded by collection order; Lane 5 - SigmaMarker S8445.*



***PICTURE N**. SDS-PAGE of the affinity purification assay of kred pET-100 TOPO 20h induction, 25ºC (second gel). Lane 1 - SigmaMarker S8445; Lane 2 – Cell extract (diluted with imidazole buffer 6 mM); Lane 3 – Flow-through.*