Molecular Engineering of Peptide Fusions with Removable Tags for Industrial Scale Protein Purification

Feasibility Study

Renato Pereira Castelão

Thesis to obtain the Master of Science Degree in

Biological Engineering

Supervisors:

Prof. Dr. Guilherme Nuno de Passos Correia Matos Ferreira

Dr. Jan-Metske van der Laan

Examination Committee

Chairperson: Prof. Dr. Duarte Miguel de França Teixeira dos Prazeres

Supervisor: Prof. Dr. Guilherme Nuno de Passos Correia Matos Ferreira

Members of the Committee: Prof. Dr. Ana Margarida Nunes da Mata Pires de Azevedo

October 2019

I have no special talent. I am only passionately curious.

Albert Einstein

ACKNOWLEDGMENTS

Looking back to these lasts 7 months, I can only say that it was such an amazing adventure, full of challenges, goals to accomplish and high expectations. This experience allowed me to acquire valuable knowledge that will, surely, contribute to my personal and professional future life. On the other hand, going through such a challenging experience was only possible due to an amazing group of people that was always by my side.

It all started with Prof. Guilherme Ferreira who helped me establishing the connection between Técnico and DSM Food Specialties, also accepting the supervision of my project. To him, and also to DSM, I would like to express my sincere gratitude for the great opportunity to develop my work with so good conditions.

After starting this internship, I had the opportunity to meet Jan-Metske van der Laan, my internal coordinator, and a scientist with so much experience in this working field. I would really like to thank him for all the valuable input and advices, for his availability when I needed to discuss results, and for all the important knowledge that he was always willing to share.

I must similarly extend my acknowledgments to Ton Verkaik. Firstly, for all the lab coaching and support converting scientific ideas into proper laboratory experiments. His expertise, along with all the recommendations and advices were of extremely importance for the proper development of this project. Secondly, for all the conversations about football, for all the recommendations about the best places to visit in The Netherlands, and for such an amazing tour in Rotterdam.

As it is impossible to mention everyone, I would also like to thank to ANA-BIO group, especially for all the support in the laboratory, from helping me to find the right materials to teaching me how to work with specific equipment.

To Catarina Neves, a very special thank you! Thank you for all the guidance and for all the discussions we had about this project even outside DSM. But even more important, thank you for the whole support during these seven months: for giving me a place to stay during my first week in Delft, for all the breakfasts and dinners, and for introducing me to a group of seven incredible persons.

Madrinha Diva, Lenny, Ticha, Pedro, Chuck, Raquel and Parreira were indeed my family in Delft, and so I would like to truly thank them for all the amazing moments we had. All of them, each one in its own way, had a very positive impact in my life, and I am grateful to call them “friends”.

To all these people, even if our paths follow different directions, I really hope to keep in touch with you and to meet you either in Portugal, or all around the world.

Even though these last seven months were critical for the conclusion of this project and will be joyfully remembered, I cannot exclude the importance of these last five years in my life.

Firstly, I address my gratitude to Instituto Superior Técnico, including all the professors, for everything that I have learned and that I will definitely take with me in my future life. Técnico also gave me a fantastic

group of friends that I will never forget. Naming all would leave me no more space for further acknowledgments, but I thank every single one for all the joyful moments we had. When tough times come, you know who your true friends really are, and fortunately I can say that I do have true friends. So, for their crucial role in my academic life, I thank Catarina Soares, Beatriz Vivar, Sofia Navalho, Guilherme Fernandes and Rafaela Pereira from the bottom of my heart. Also, I cannot forget to thank to my special Erasmus friends, António Fragoso and Rosa Lourenço, for all the moments in Italy, for the King's Day, and for all the moments that are still to come.

(The following paragraph is written in Portuguese, considering the addressed persons.)

Finalmente, resta-me agradecer a toda a minha família que sempre me apoiou neste meu percurso. Particularmente, quero agradecer do fundo do coração às três pessoas mais importantes da minha vida. Ao meu Pai e à minha Mãe que sempre me apoiaram em todas as minhas decisões, mesmo quando tais implicaram ir viver inicialmente para Lisboa, depois para Milão e, por fim, para Delft. Por terem estado sempre ao meu lado ao longo deste percurso, quero expressar o meu maior obrigado. Por fim, agradecer ao meu Irmão que, mesmo sendo tão novo, sempre conseguiu perceber as minhas motivações e dar-me a força que sempre precisei. Tudo o que consegui alcançar até hoje, é graças a vocês!

"Lots of people want to ride with you in the limo, but what you want is someone who will take the bus with you when the limo breaks down". To my dear friends and family, thank you for taking buses with me!

Obrigado.

Grazie.

Bedankt.

Thank you.

ABSTRACT

Food processing industry is experiencing major contributions from enzyme technology in several applications and, consequently, the associated downstream processing is one of the areas at the forefront of enzyme research and development. Affinity chromatography is the most specific type of chromatography enabling the achievement of highly purified proteins in a single-step purification. The target protein is purified with a linked affinity tag that is usually subsequently removed with specific proteases. However, this extra proteolytic step usually raises a few difficulties turning the process more complex and costly.

The present project focused on a feasibility study of an affinity chromatography system using inteins as self-removable tags. Two target proteins, two inteins (*Mxe* GyrA and *Mtu* RecA) and two affinity tags (chitin-binding domain (CBD) and polyhistidine tag) were initially tested in six different fusion constructs.

The most promising purification was performed with a fusion expressing the *Mtu* RecA intein and the CBD as the affinity tag: using around 5% of the column capacity, it was possible to purify the target protein with an estimated purity higher than 95% and a purification yield of 96%. However, when reproducing the same purification experiment using about 70% of the affinity column capacity the purification yield dropped to 3%, indicating a possible unsuccessful intein cleavage.

This feasibility study allowed to conclude that this intein-mediated purification system may be valuable for DSM Food Specialties, but further optimization of the chromatography and intein cleavage is still needed.

Keywords: Protein purification; Affinity chromatography; Intein; Chitin; Fusion protein; Self-cleaving tag.

RESUMO

A tecnologia enzimática tem fornecido contribuições significativas a diversas aplicações da indústria alimentar. Consequentemente, o *downstream processing* de enzimas tem sido uma das áreas alvo mais importantes de investigação e desenvolvimento. A cromatografia de afinidade é o tipo de cromatografia mais específico, que permite a obtenção de elevados graus de pureza num único passo de purificação. A proteína alvo é purificada com uma *tag* de afinidade que, tipicamente, é posteriormente removida com recurso a proteases. No entanto, este passo proteolítico adicional costuma levantar algumas dificuldades, tornando o processo mais complexo e dispendioso.

O presente projeto assenta num teste de exequibilidade ao nível de uma cromatografia de afinidade usando inteínas como *self-removable tags*. Duas proteínas alvo, duas inteínas (*Mxe* GyrA e *Mtu* RecA) e duas *tags* de afinidade (*chitin-binding domain* (CBD) e *polyhistidine tag*) foram inicialmente testadas em seis diferentes proteínas de fusão.

A purificação mais promissora foi realizada com uma fusão a expressar tanto a inteína *Mtu* RecA como a *tag* de afinidade CBD: usando aproximadamente 5% da capacidade da coluna cromatográfica, foi possível purificar a proteína alvo com uma pureza estimada superior a 95% e com um rendimento de purificação de 96%. No entanto, quando reproduzindo a mesma purificação usando cerca de 70% da capacidade da coluna, o rendimento de purificação foi reduzido para 3%, indicando uma falha na clivagem da inteína.

Este teste de exequibilidade permitiu concluir que este sistema de purificação mediado por inteínas poderá adicionar valor à DSM Food Specialties. No entanto, serão necessárias otimizações adicionais ao nível da cromatografia e da clivagem da inteína.

Palavras-chave: Purificação de proteínas; Cromatografia de afinidade; Inteína; Quitina; Proteína de fusão; *Self-cleaving tag*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	Absorbance
AC	Affinity Chromatography
CBD	Chitin-Binding Domain
CTC	Construct
CV	Column Volume
Da	Dalton
DBC	Dynamic Binding Capacity
DSP	Downstream Processing
His-tag	Polyhistidine-tag
***Mtu* RecA**	*Mycobacterium tuberculosis* RecA
MW	Molecular Weight
***Mxe* GyrA**	*Mycobacterium xenopi* Gyrase A
OD	Optical Density
PAGE	Polyacrylamide-gel electrophoresis
pI	Isoelectric point
POI	Protein of Interest
rpm	Revolution per minute
RT	Room Temperature
RTX	Repeat-in-Toxin
SBC	Static Binding Capacity
w/v	Weight/Volume

1. INTRODUCTION

1.1. Enzymes: an overview

In order for a reaction to occur, the reactant molecules must contain enough energy to cross the potential energy barrier to form the products. This energy is termed activation energy and can be defined as the amount of energy required to bring all molecules in 1 mole of a substance at a given temperature to the transition state at the top of the energy barrier. At this barrier, there is equal probability for them to undergo a reaction to form products or fall back into the unreacted molecules. Therefore, the activation energy represents a barrier which must be overcome before the reaction can proceed. All molecules possess varying amounts of energy, but generally, only a few have enough energy for reaction. The lower the potential energy barrier, the more reactants have enough energy for reaction and, thus, the faster the reaction can occur. All catalysts, including enzymes, accelerate chemical reactions by finding a lower pass over the energy barrier. In other words, the catalytic power of enzymes relies on their ability to stabilize the transition state of a reaction relative to that of the ground state ($\Delta G_{ETS*} > \Delta G_{ES}$), decreasing the Gibbs energy of activation (standard free energy of activation), when comparing to the uncatalyzed reaction. This is represented in Figure 1, as well as the fact that the enzyme-substrate complex is stabilized relative to the free species in solution [1, 2].

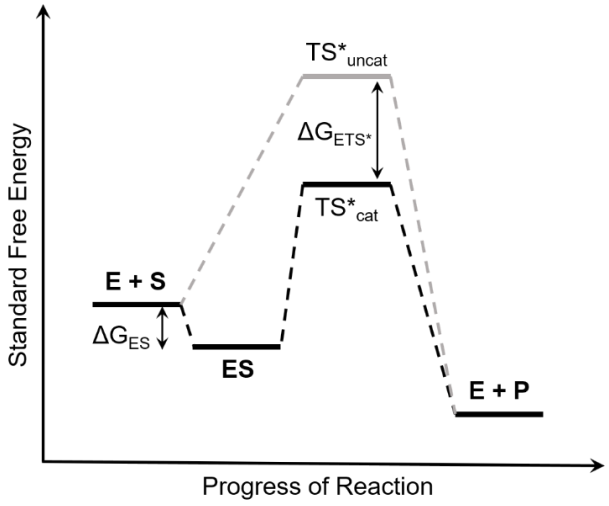

Figure 1 Energy diagram showing comparative free energy changes (ΔG) for catalyzed (cat) and uncatalyzed (uncat) reactions. S, substrate; E, enzyme; P, product; ES, enzyme-substrate complex; TS*, transition state; ΔG, standard free energy of activation (adapted from [1]).

Like all catalysts, enzymes increase the rate of chemical reactions without altering the equilibrium constant of the catalyzed reaction and without undergoing any net change in their structure. Enzymes are very efficient catalysts, often far superior to conventional chemical catalysts, due to their outstanding specificity and selectivity. It is this specificity that draws the interest of technologists and scientists for a wide range of applications, from industrial processes to clinical diagnosis or from environmental remediation to drug or food formulation as obvious examples of the key role that enzymes play in our daily life [1].

1.1.1. Structure

Enzymes are proteins, and so are made up of amino acids linked together by amide bonds. Each peptide chain has one free amino end, the N-terminus, and one free carboxyl end, the C-terminus. Since enzymes are proteins, then they can be structurally analyzed taking advantage of numerous methods of protein chemistry, molecular biology and molecular biophysics [1, 3].

The sequence of amino acids of an enzyme determines its primary structure. Due to the properties of the amino acids, enzymes may possess a helical structure (α-helices) and other possible arrangements such as turns and pleated sheets (β-sheets) resulting in the secondary structure. The tertiary structure of an enzyme refers to its three-dimensional (3D) arrangement in order to minimize non-favorable thermodynamic interactions with its environment. This 3D structure forms sites that are directly involved in binding the substrate, prosthetic groups and any other cofactors. These cofactors are nonprotein components required by many enzymes for their catalytic activity and can range from metal ions to larger complex organic molecules. Lastly, a lot of known enzymes are composed of two or more polypeptide chains, usually linked to each other by noncovalent interactions. These polypeptide chains are termed subunits and may be identical or different from each other. The quaternary structure is the combination and spatial arrangement of these subunits [1].

It is also possible to distinguish between binding and catalytic sites. Binding sites, in the binding domain of the protein, link to specific groups in the substrate, ensuring that the enzyme and the substrate are held together in a fixed orientation with respect to each other. Catalytic sites, in the catalytic domain, link the group or groups of the substrate that are directly involved in the catalysis. Both the substrate binding site and the catalytic binding site are usually near one another and frequently overlap. The region of an enzyme 3D structure that contains the substrate and catalytic binding sites is called the active site or active center. These bindings of substrates to enzymes requires the interplay of electrostatic, hydrogen bonds and van der Waals interactions [1].

The three-dimensional structure of an enzyme can be obtained at high resolution by X-ray crystallography or NMR spectroscopy, making it possible to not only determine the detailed structures of enzymes, but also broadly understand the principles of protein structure [3].

1.1.2. Classification

According to the report of the first Enzyme Commission in 1961, enzymes fall into six general classes according to the type of reaction catalyzed (oxireductases, transferases, hydrolases, lyases, isomerases and ligases), and they can be identified through a sequence of four numbers. The first number refers to the main type of reaction. The second and the third numbers indicate the subclass and sub-subclasses, respectively, regarding the nature of the actual reaction that is catalyzed. Lastly, the fourth number is the serial number in its particular subclass, mostly related to the enzyme substrate specificity [1, 2, 3].

Oxireductases (EC1) encompasses all enzymes that catalyze redox reactions. Typical examples of these enzymes are dehydrogenases, oxidases, peroxidases and oxygenases. The systematic name is formed according to "*donor*:*acceptor* oxireductase" [1, 2, 3].

Transferases (EC2) catalyze the transfer of a specific group, such as methyl, acyl, amino, glycosyl, or phosphate, from one substance (generally regarded as the donor) to another (generally regarded as the acceptor). The systematic name is formed according to "*donor*:*acceptor group*transferase" [1, 2, 3].

Hydrolases (EC3) catalyze the hydrolytic cleavage of C-O, C-N, C-C, and some other bonds. Although the recommended name often consists simply on the substrate name with the suffix-ase, the systematic name always includes hydrolase. According to the type of bond cleaved, a distinction is made between peptidases, esterases, glycosidases, phosphatases, etc. [1, 2, 3].

Lyases (EC4) catalyse the cleavage of C-C, C-O, C-N, and other bonds by other means than by hydrolysis or oxidation (elimination reactions). A few examples are decarboxylases, aldolases and dehydratases, which are responsible for elimination of CO_2, aldehyde and water, respectively. The systematic name is formed according to "*substrate group*-lyase" [1, 2, 3].

Isomerases (EC5) catalyze geometric or structural rearrangements within a molecule. Depending on the different possible types of isomerism, there are racemases, epimerases, *cis-trans*-isomerases, tautomerases, mutases or cycloisomerases [1, 2, 3].

Lastly, **ligases** (EC6), also called synthetases, are known to catalyze the joining of two molecules, coupled with hydrolysis of a pyrophosphate bond in ATP or a similar triphosphate. The systematic name is formed according to "*X*:*Y* ligase" [1, 2, 3].

1.1.3. Enzymes in Food Industry

Enzymes have several distinct advantages over conventional chemical catalysts and find a wide range of applications in different bioprocesses. There are many important areas in which enzyme technologists are currently applying their art and which are at the forefront of enzyme research and development. Food processing industry has experienced major contributions from enzyme technology in several applications including bioconversions and synthesis, improvement of extractions, reduction of viscosity, changes in functionality and flavor modification. The advances in genetic engineering have been allowing the production of proteins from a specific organism (sometimes pathogenic or unculturable) in different hosts. For that reason, when choosing the production strain, two aspects should be considered: ideally the enzyme should be secreted from the cell, making the recovery and purification process much simpler, and the organism should be able to produce a high amount of the desired enzyme in a reasonable time frame. Also, note that for food applications, the enzyme must be obtained from sources which have GRAS (generally regarded as safe) status [2, 4].

Table 1 summarizes some important enzymes along with their applications in five different industries.

Table 1 An overview of enzymes used in the industrial segments of food and feed processing with respective application (adapted from [2, 5, 6]).

Industry	Enzyme	Class	Application
Dairy	Catalase	Oxidoreductases (EC1)	Milk sterilization, peroxide decomposition
	Esterase	Hydrolases (EC3)	Accelerated cheese ripening, milk fat treatment
	Lactase		Hydrolysis of lactose to produce low lactose/lactose free milk
	Lipase		Aging, ripening and general flavor characteristics
	Phosphatase		Monitoring of pasteurization
	Protease		Milk clotting, flavor improvement
Baking	Lipoxygenase	Oxidoreductases (EC1)	Dough strengthening, bread whitening
	Amylase	Hydrolases (EC3)	Dough conditioning, bread softness and volume, ensure uniform yeast fermentation, improved machinability, increase shelf life
	Xylanase		Dough conditioning
	Proteases		Improved machinability
	Lipase		Dough stability and conditioning (*in situ* emulsifier)
	Phospholipase		Dough stability and conditioning (*in situ* emulsifier), improved quality of bread
Brewing	Glucose oxidase	Oxidoreductases (EC1)	Removal of glucose and oxygen from different food products, flavor and color preservation
	Laccase		Flavor enhancer
	Amylase	Hydrolases (EC3)	Starch conversion to maltose for fermentation, low calorie beer
	Glucanase		Viscosity reduction of wort or beer
	Cellulase		Hydrolysis of complex carbohydrate cell walls
	Proteases		Chill proofing (prevention of chill haze), improved shelf life of beers
	Acetolactate decarboxylase	Lyases (EC4)	Beer maturation
	Glucose isomerase	Isomerases (EC5)	Low calorie beer
Meat	Protease	Hydrolases (EC3)	Meat tenderizer
Animal feed	Phytase	Hydrolases (EC3)	Digestibility
	Xylanase		
	Glucanase		

1.1.3.1. Market and Future Prospects

Enzyme based strategies are currently being increasingly used for their multifarious applications in foods, dairy, pharmaceutical and beverage industries and, consequently, are replacing traditional chemical methods in both laboratories, and industries. Enzymes are extensively used in food processing mainly in United States, Europe, and Japan [2]. In fact, major enzyme producers are located in Europe, USA and Japan, with well-known companies such as Novozymes, Danisco, Genencor, DSM and BASF. Out of the total industrial demand for enzymes, in terms of end-use, food and feed represent the largest segment [4], essentially with starch and sugar processing, dairy, and bakery applications [2].

Regarding the global market for the industrial enzymes, a report published by BBC Research in 2014 [7] states that the global market for industrial enzymes was worth nearly $4.5 billion in 2012 and nearly $4.8 billion in 2013. A more recent report from 2018 [8], also published by BBC Research, declares that global industrial enzyme market should reach $7.0 billion by 2023 from $5.5 billion in 2018.

The major barrier in the use of enzymes at large scale is their unstable nature, cost of production, and non-availability of specific enzymes to work under the industrial process conditions (temperature, pH, ionic strength and organic solvents). For this reason, to really achieve the full potential of enzymes in food industry, these technological and implementation barriers need to be overcome [2].

Lastly, for realizing the actual potential of enzymes in industrial application, there are still plenty of improvements to achieve regarding the development of innovative and sustainable process technologies. This include the development of effective tools and techniques for rapid screening of biocatalysts with desired functions and properties; the improvement of protein engineering techniques based on site specific or random mutagenesis, and directed enzyme evolution to tailor the enzyme for a specific biotransformation; and the development of efficient high-expression hosts (including genetically modified microorganisms) and suitable cultivation technology to improve the cost-effective production of enzymes [2]. It is important to understand that the enzyme industry is an excellent position to contribute to a cleaner environment. Firstly, enzyme technology offers industries and consumers an opportunity to replace processes using aggressive chemicals with mild, non-toxic enzyme processes. And secondly, the biotechnological processes used to produce enzymes have been showing to have a minimal impact on the environment [4].

1.2. Catalytic Activity of Enzymes

1.2.1. Enzyme Activity Assays

1.2.1.1. Measuring Catalytic Activity using Reaction Rate

The conversion rate of substrates, *v*, can be measured to determine the catalytic activity. Considering Michaelis-Menten kinetics, if the substrate concentration is kept at a level considerably above Michaelis constant ([S] >> K_M), then the enzyme is saturated with substrate and the reaction proceeds with the maximum rate (zero order reaction). In this condition, the rate of the enzymatic reaction is proportional

to the concentration of the enzyme, being this situation attempted to be reached in enzyme activity measurements [3, 9].

1.2.1.2. Definition of Units

In 1961, the Enzyme Commission of the International Union of Biochemistry defined the International Unit, U, as the amount of enzyme which, under optimized standard conditions, catalyzes the conversion of 1 µmol of substrate per minute. With respect to SI basic units, the Expert Panel on Quantities and Units (EPQU) of the International Federation of Clinical Chemists (IFCC) and the Commission on Quantities and Units in Clinical Chemistry (CQUCC) of IUPAC defined the base unit *katal* as the catalytic amount of any enzyme that catalyzes a reaction rate of 1 mol of substrate per second in an assay system. However, this last system unit is not commonly used. All assay conditions, namely the temperature, must be stated, being precisely controlled and kept constant to achieve reproducible results [1, 3].

However, for practical reasons, many enzyme reactions cannot be monitored by measuring the stoichiometric amount of substrate consumed or product formed, and for that reason, the catalytic activity of a particular enzyme might be impossible to express in International Units. As an example, the unit of a restriction endonuclease is defined as the catalytic activity of the enzyme that yields a typical cleavage pattern, detectable after electrophoresis, with a precise amount (usually 1 µg) of a particular DNA under defined incubation conditions. Other parameters that can be used to define the activity of such enzymes are the degradation of a nucleic acid or the incorporation of nucleotides into a nucleic acid, expressed in micrograms, nanomoles, absorbance units, or number of base pairs. Also, for practical reasons, different time periods, such as 1, 10, 30 or 60 minutes, may be chosen as references [3].

1.2.1.3. Measuring Methods

Nowadays, photometry is one of the preferred methods of enzyme assay and can be carried out most quickly and convenient when the substrate or the product is colored or absorbs light in the ultraviolet or visible (UV/Vis) range, since the rate of appearance or disappearance of a light-absorbing product or substrate can be followed with a spectrophotometer. For this method, the Lambert-Beer law, which defines a relationship between absorbance and concentration, is usually followed [3].

However, there are other (less used) methods that can also be applied to determine the catalytic activity of raw or purified enzyme preparations, even with some associated disadvantages. Between these methods, it is possible to mention luminometry (using bioluminescence), radiometry (when radioactively labeled substrates are used), potentiometry (to measure reactions in which protons are produced or consumed), calorimetry (for enzymatic reactions that evolve heat), viscosimetry (in which activity is determined by the change of viscosity per unit time), or electrophoresis (to determine the catalytic activity of nucleases, especially restriction endonucleases – usually, it is estimated the minimum amount of enzyme which completely converts a given substrate to the typical fragment pattern for that enzyme) [3].

1.2.2. Quality Evaluation of Enzyme Samples

The quality of an enzyme sample is considered to be function of its activity, purity and stability.

1.2.2.1. Specific Activity

One of the most significant quality criteria of an enzyme sample is its specific activity, defined as the catalytic activity related to its total protein content. Specific activity is usually expressed as units per milligram of enzyme. If the molecular mass of the enzyme is accurately known, it is possible to express the activity as the molar catalytic activity, defined as the number of units per millimole of enzyme. This, however, may not correspond to the conversion per enzyme active site, since the enzyme may contain more than one active site. If the number of active sites per mole is known, then the activity may be expressed as the catalytic center activity [1, 3].

1.2.2.2. Contaminating Activities

The content of contaminating activities, obtained by the presence of other enzymes in the original material, is also an important quality criterion for enzymes. This is usually related to the activity of the main enzyme and expressed as a percentage [3]. Depending on the specific application of an enzyme, the activity of different impurities should be determined, to guarantee that they do not have a negative impact on the final application.

1.2.2.3. Electrophoretic Purity

Electrophoresis is imperative in the evaluation of purity due to its sensitivity, despite being far inferior in the determination of contaminating activities. Furthermore, enzymatically inactive proteins usually do not interfere with enzymatic analysis. Electrophoresis reveals to be a very important method, for instance, when isoenzymes need to be considered, since they cannot be distinguished by their catalytic function, but only by physical properties, such as electric charge [3].

In addition to the usual electrophoresis in polyacrylamide gels under denaturing and reducing conditions (SDS-PAGE), where proteins are primarily separated by their mass, electrophoresis with 2-dimensional gels (2D-PAGE) and isoelectric focusing (IEF) are also commonly used. In 2D-PAGE, proteins are separated by their isoelectric point (pI) in the first dimension, and by their relative molecular weight in the second dimension. On the other hand, IEF, an electrophoretic procedure capable of extremely high resolving power, separates proteins according to their isoelectric point (pI) [10, 11].

1.3. Downstream Processing of Proteins

The purification of a product, the so-called downstream process (DSP), tends to be one of the costliest aspects of modern bioprocessing, especially for proteins. Different downstream processing methods for protein purification can be performed, from precipitation, crystallization, membrane-based separation or chromatography [2, 3, 12]. However, chromatography is still the major tool on all levels of DSP, from the first capture to the final polishing step. It is also important to understand that, depending on the type of

final application, partially purified enzyme preparations may be sufficient, or there might be the need to have highly purified enzymes. For this reason, the final desired purity is a determining factor when choosing a right DSP method. Furthermore, it is important to keep in mind that if several purification steps are required to reach a high level of purity, the yield in each step is critical because even a small product loss in each purification step adds up to significant losses over the entire downstream process [3, 13, 14].

1.3.1. Chromatography

1.3.1.1. Principle of Working

Chromatography is the science which studies the separation of molecules based on differences in their structure and/or composition, being based on the principle where molecules in a mixture applied into a stationary phase are separated from each other while moving with the aid of a mobile phase. Thus, this type DSP comprises a wide variety of techniques characterized by an ability to separate complex compound mixtures on the basis of small differences in physiochemical parameters, such as, size and shape, net charge, hydrophobicity, isoelectric point or biological function. Because of these differences, some components of the mixture stay longer in the stationary phase in the chromatography system, while others pass rapidly into mobile phase, and leave the system faster. Based on this approach, three main components form the basis of the chromatography technique: the stationary phase (a solid phase, or a layer of a liquid adsorbed on the surface of a solid support), a mobile phase (composed of liquid or gaseous component) and the molecules to be purified [2, 14, 15].

The most common physical configuration is column chromatography, in which the stationary phase is packed into a column, through which the mobile phase (the eluent) is pumped. The sample to be separated is introduced into one end of the column and its various components travel with different velocities through the column and are subsequently detected and collected at the other end [16].

In the context of the present project, the mobile phase is a liquid, most often an aqueous buffer. Consequently, these techniques are versions of liquid chromatography. Only for the separation of more volatile compounds a gaseous mobile phase is used, being this gas chromatography technique extremely powerful and also widely applied [16].

Chromatography is of fundamental importance to enzyme purification, having the great advantage that it can be performed under physiological conditions (aqueous solution, temperature between 4 and 40°C, pressure below 10 bar, neutral pH), being presently the only cost-competitive, scalable, high-resolution method for protein isolation and separation [13].

1.3.1.2. Types of Chromatography

It is possible to identify several types of chromatography, each one allowing protein separation according to different molecular properties. Nevertheless, the most commonly used chromatography types are Ion-Exchange Chromatography (IEX), Size Exclusion Chromatography (SEC) and Affinity Chromatography (AC).

Ion-exchange chromatography (IEX) is a separation technique based on the charge of protein molecules. Enzyme molecules vary considerably in their charge properties and will exhibit different degrees of interaction with a charged chromatography media according to differences in their overall charge, charge density, and surface charge distribution. Since all molecules with ionizable groups can be titrated, their net surface charge is highly pH dependent. This means that the net surface charge of proteins, which are built up of many different amino acids containing weak acidic and basic groups, will change gradually as the pH of the environment changes. In fact, each protein has its own unique net charge *versus* pH relationship which can be visualized as a titration curve, reflecting how the overall net charge of the protein changes according to the pH of the surroundings (Figure 2). This property is then used to separate proteins by chromatography on anion exchangers (positively charged resin) or cation exchangers (negatively charged resin). In an IEX separation, reversible interactions between charged molecules and oppositely charged IEX media are controlled in order to favor binding or elution of specific molecules achieving, that way, the desired separation. A protein that has no net charge at a certain pH, equivalent to its isoelectric point (pI), will have its interactions with a charged medium minimized, and will be eluted from the column. However, at a pH above its pI, a protein will bind to a positively charged medium (anion exchanger) and, at a pH below its pI, a protein will bind to a negatively charged medium (cation exchanger). The type of charged group determines the type and strength of the ion-exchanger, while their total number and availability determines the capacity [2, 3, 17].

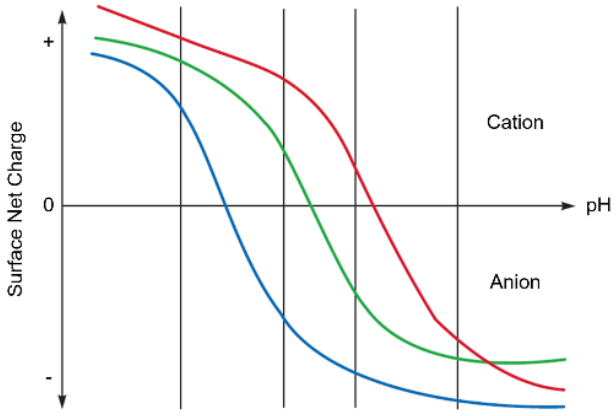

Figure 2 Theoretical protein titration curves, showing how net surface charge varies with pH (adapted from [17]).

In **size exclusion chromatography** (SEC), also called gel chromatography or gel filtration, molecules are separated according to their size and shape. Molecules larger than the largest pores in the gel beads or, in other words, above the exclusion limit, pass straight through the column without entering the pores and are eluted first. Smaller molecules, on the other hand, penetrate through the open network in the pores of the gels to a varying degree depending on their size and shape, have their passage retarded through the column and are eluted in order of decreasing molecular mass. By varying the degree of cross-linking, gels of different porosities and with different fractionation ranges can be obtained. Note, however, that for this type of chromatography, concentrated solutions are necessary for separation, since the sample volume that can be applied to a column is limited to around 10% of the column volume [2, 3].

Affinity chromatography separates proteins based on a reversible interaction between a protein and a specific ligand coupled to a chromatography matrix. Along the present project, all the main purification procedures were performed based on affinity chromatography. Hence, a more detailed theoretical analysis on this type of chromatography is made in the following section 1.3.1.3.

1.3.1.3. Affinity Chromatography: a theoretical background

Overview [2, 3, 18, 19]

Affinity chromatography is the most specific type of chromatography, which separates proteins based on a reversible interaction between a protein (or a group of proteins) and a specific ligand coupled to a chromatography matrix. This technique usually offers high selectivity, high resolution, and usually high capacity for the protein of interest. Affinity chromatography is unique in purification technology since it is the only technique that enables the purification of a biomolecule on the basis of its biological function or individual chemical structure. Suitable ligands can be substrate analogues, enzyme inhibitors, dyes, metal chelates or, for instance, antibodies. When an unpurified enzyme mixture is passed down a column containing the affinity matrix, only the desired enzyme is retained, and other proteins are washed away.

Immunoaffinity chromatography, when monoclonal antibodies are used as ligands, occupies a unique place in purification technology. In this procedure, enzymes can be purified by immobilizing antibodies specific for the desired enzyme. A more general method offers the synthesis of a fusion protein with protein A tag, a *Staphylococcus* protein with a high affinity for many immunoglobulins, especially for the IgG class of antibodies. In this way, enzymes that usually do not bind to an antibody can be purified by immunoaffinity chromatography.

Immobilized metal-affinity chromatography (IMAC) is also widely employed to purify recombinant proteins containing a short affinity tag consisting of polyhistidine residues. IMAC is based on the interactions between a transition metal ion (Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}) immobilized on a matrix and specific amino acid side chains. In fact, histidine is the amino acid that exhibits the strongest interaction with these immobilized metal ion matrices, as electron donor groups on the histidine imidazole ring readily form coordination bonds with the immobilized transition metal. Because of this, peptides containing sequences of consecutive histidine residues are efficiently retained on IMAC column matrices.

To sum up, affinity chromatography reveals to be very important since its high selectivity enables many separations to be achieved in one simple step, including, for example, purification of monoclonal antibodies or fusion proteins.

Media

Regarding the affinity resin, three typical terms are commonly used: the matrix (should be chemically and physically inert and it is the support for ligand attachment), the spacer arm (used to improve binding between ligand and target molecule by overcoming any effects of steric hindrance) and the ligand (molecule that binds reversibly to a specific target molecule or group of target molecules) [3, 18].

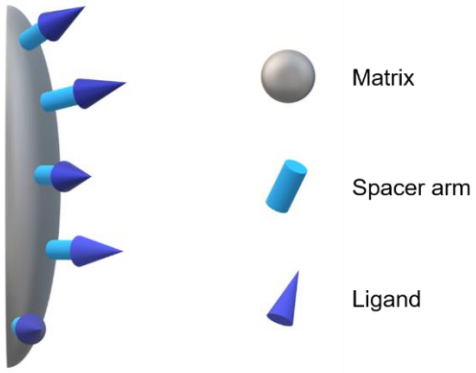

Figure 3 Representation of a ligand covalently bound to the chromatography matrix through a spacer arm.

Successful affinity purification requires a biospecific ligand that can be covalently attached to a chromatography matrix. The coupled ligand must retain its specific binding affinity for the target molecules and, after washing unbound material, the binding between the ligand and target molecule must be reversible to allow the target molecules to be removed in an active form. Note, however, that this type of chromatography can also be used to remove specific contaminants from a protein sample. Table 2 summarizes a few typical biological interactions, frequently used in this type of chromatography. The ligand is immobilized on the solid phase, while the counterligand (most often a protein) is adsorbed from the extract that is passing through the column [3, 18].

Table 2 Examples of biological interactions used in affinity chromatography (adapted from [20]).

Ligand	Counterligand
Antibody	Antigen, virus
Inhibitor, substrate analogue	Enzyme
Lectin	Polysaccharide, glycoprotein, cell surface receptor, membrane protein
Nucleic acid	Nucleic acid-binding protein (enzyme or histone), complementary base sequence
Hormone, vitamin	Receptor, carrier protein
Metal ions	Poly His fusion proteins, native proteins with histidine, cysteine and/or tryptophan residues on their surfaces
Glutathione	Glutathione-S-transferase (GST) or GST fusion proteins

For the selection of the right media, a ligand already coupled to a matrix is the simplest solution. If a ligand is available, but there is not any available media using it, then it firstly needs to be coupled to a pre-activated matrix, a matrix which have been chemically modified to facilitate the coupling of specific types of ligands. Literature comprises several experimental details of commonly used procedures for the immobilization of different ligands to various matrices, from beaded agarose to other polyol matrices [18, 20]. Lastly, if there is no suitable ligand available, then it should be considered whether it is worth

the time and effort involved to obtain a ligand and to develop a specific affinity medium, or if it is more convenient to use alternative purification techniques.

Incorporation of Tags

Affinity tags are commonly placed on either the N or the C-terminus of recombinant proteins. Optimal placement of the tag is, however, protein specific. A potential problem is inaccessibility of the protein tag to the specific ligand present in the chromatography matrix due to occlusion of the tag in the folded protein. In principle, depending on the tag size and charge, it is also possible that the affinity tag may interfere with protein activity. For this reason, it might be useful to remove the affinity tag after purification by use of a protease cleavage site inserted between the tag and the protein [19].

Mode of Operation

Before starting the chromatography, several procedures should be performed. Firstly, samples should be prepared so that they are clear and free from particulate matter, since this will avoid the clogging of the column, may reduce the need for stringent washing conditions and will extend the life of the chromatographic medium. Also, during the preparation of the samples, all the components known to interfere with the binding between the target molecule and the ligand should be removed. The affinity of the ligand should also be tested: very low affinity will result in poor yields due to the fact that the target protein may just be washed from the column during sample application. On the other hand, too high affinity may also result in low yields, since the target molecule may not dissociate from the ligand during elution. The flowrate that gives the most efficient binding during sample application should also be tested, since this parameter can vary according to not only the specific interaction between the target protein and the ligand, but also their concentrations. Regarding this topic, for weak affinity interactions (between the target molecule and the ligand) samples should be usually loaded at a lower flow rate, while for interactions with a strong affinity that quickly reach an equilibrium, a higher flow rate can be used [18].

When starting the affinity chromatography, the column must be submitted to a first **pre-equilibration** step with binding buffer (mobile phase), so that the stationary phase can be equilibrated to the desired conditions. These conditions, essentially established by the binding buffer, should be selected in order to ensure not only optimal binding conditions, but also optimal stability conditions to the protein of interest. For these reasons, several important factors dictate the choice of the mobile phase buffer including, for instance, buffer strength, pH and presence of ions known to stabilize the protein of interest, always guaranteeing, on the other hand, that it has no affinity to the stationary phase [18].

After equilibrating the column, the feedstock containing the protein of interest and all the other remaining proteins can be loaded into the column – **sample loading**. It is important to make sure that all unbound material is washed through the column by the binding buffer (or, if needed, a washing buffer) by a **washing** step after sample loading. The chromatography is usually monitored by UV absorbance at 280 nm, and the washing step is finished when the original base line is reached. This step is critical to improve the purity of the eluted target substance. If nonspecific binding of untagged proteins or nonspecific hydrophobic interactions are suspected, copurification with the desired protein may occur

and, for that reason, stringent washing conditions can be used in this step if the protein of interest is known to have a relatively strong affinity to the stationary phase. Including low levels (up to 1%) of the nonionic detergent Triton X-100 or Tween 20, addition of salt (up to 500 mM NaCl), glycerol (up to 20%) or low levels of ethanol (up to 20%) in protein buffers may reduce these nonspecific interactions with the matrix. Alternatively, for some affinity purifications such as IMAC, a washing buffer with a pH lower than that of the binding buffer can be employed to remove nonspecifically bound proteins [18, 19, 20].

After the whole sample has been loaded and the column washed, all conditions are met to begin the **elution** phase. There is no generally applicable elution scheme for all affinity media, and selective or non-selective methods may be applied. Selective elution methods are applied in combination with group-specific ligands, whereas non-selective elution methods are used in combination with highly specific ligands. Ligand-protein interaction is often based on a combination of electrostatic, hydrophobic and hydrogen bonds, so agents that weaken this kind of interactions may be expected to function as efficient eluting agents. Four different elution methods can be distinguished, as shown in Figure 4. Method 1 is the simplest case, in which a change of buffer composition elutes the bound substance without harming either it or the ligand. For method 2, extreme pH or high concentrations of chaotropic agents are applied for elution, but these may cause permanent or temporary damage both to the ligands and to the target protein. Lastly, in both methods 3 and 4, specific elution is performed by adding a substance that competes for binding either to the ligand (method 3) or to the target protein (method 4). A compromise may have to be made between the harshness of the eluent required for elution and the risk of denaturing the eluted material or damaging the ligand on the affinity medium [18, 20].

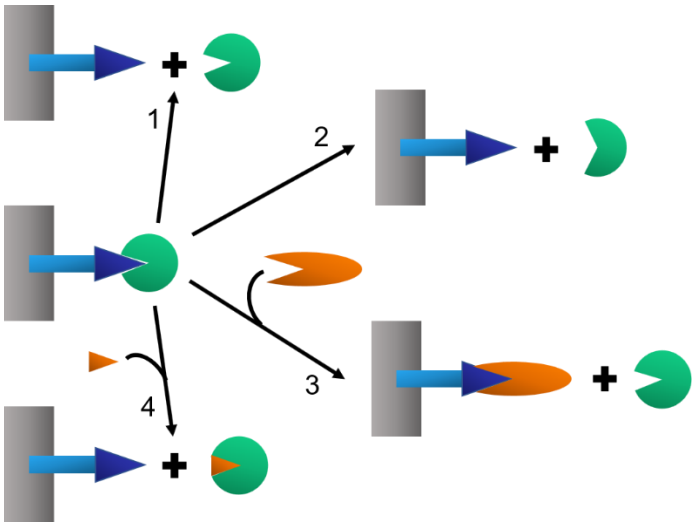

Figure 4 Scheme of four different elution methods (adapted from [18]).

General elution methods to reverse the interaction between the ligand and the target protein of the specific interaction may be applied in a gradient or step conditions and include pH elution, ionic strength elution, competitive elution, reduced polarity of eluent and chaotropic eluents. Briefly, pH elution is based on a change in pH that alters the degree of ionization of charged groups on the ligand and/or the bound protein. The exact mechanism for elution by changing the ionic strength depend upon the specific interaction between the ligand and the target protein (usually this method elutes proteins bound by

predominantly electrostatic interactions), but this mild elution may be performed using a buffer with increased ionic strength, usually up to 1 M NaCl (although occasionally 2 or 3 M salt may be required). In competitive elution, selective eluents that compete either for binding to the target protein or to the ligand are used to separate substances on a group specific medium or when the binding affinity of the ligand/target protein interaction is relatively high. Reducing the polarity of the eluent (typically dioxane up to 10% or ethylene glycol up to 50%) may also promote elution without inactivating the eluted substances. Lastly, if the previous elution methods fail, deforming buffers with chaotropic agents such as guanidine hydrochloride or urea (moderate concentrations 4-6 M) or chaotropic salts such as potassium thiocyanate (KSCN), potassium cyanate (KCNO) and potassium iodide (KI) (in the concentration range 1-3 M), known to alter the structure of proteins, can be used [18, 20].

Broadening of adsorption peaks is often a problem, that can be due to slow diffusion mass transport, slow equilibration kinetics or a wide range of binding affinities in the system. Peak shapes are often improved by reversing the direction of flow during desorption, as the distance in the column covered by the desorbed protein is minimized – reversed elution [20].

After elution, a **regeneration** step might be performed in order to remove any molecules still bound to the adsorbent, ensuring that the full capacity of the column is available for the next run. The exact extent of reuse depends on the nature of the sample and the stability of the ligand and matrix with respect to the elution and cleaning conditions used. The type of regeneration strongly depends on the used matrix and ligand. This step may be performed using, for instance, a high ionic strength buffer, inducing a change in the pH or even include detergents and denaturing agents, depending on the stability of the ligand. Lastly, the column should be **re-equilibrated** with several column volumes of starting buffer to be ready for the next run [20].

Dynamic Binding Capacity [16, 17, 21, 22, 23]

The capacity of an affinity medium is a quantitative measure of its ability to adsorb target molecules, determining how much resin is needed in order to purify a certain amount of protein. The capacity of a chromatography medium can be measured in two ways: static binding capacity (SBC) and dynamic binding capacity (DBC). SBC, normally measured in batch mode, is the maximum amount of protein that can bind to a chromatography resin at given conditions, often obtained during excess load of sample. On the other hand, DBC is measured under operating conditions i.e., in a packed affinity chromatography column during sample application at a given flow rate and is the amount of protein that binds before a significant breakthrough of the target protein occurs. DBC is determined by loading a sample containing a known concentration of the target protein and monitoring the flow through. The protein will then bind to the resin to a certain breakpoint before unbound protein will flow through the column.

For these reasons, it is possible to conclude that the dynamic binding capacity is a more representative protein capacity of the resin, since it is determined under the same running conditions as a purification run, whereas the static binding capacity is typically determined with a high protein loss and not under typical running conditions.

A breakthrough curve is generated by graphing the amount of protein loaded *versus* the percent breakthrough. Because the DBC is measured under operating conditions, this value can be used to infer about the maximum amount of target protein that can be loaded into the column in order to avoid unnecessary losses. In contrast, although SBC reports the maximum protein quantity that a resin can bind, it does not state the amount of protein that failed to bind under these conditions. Note that, especially for affinity chromatography, the binding equilibrium may be slow, often not resulting in a sharp front at the outlet, but rather a more rounded shaped front.

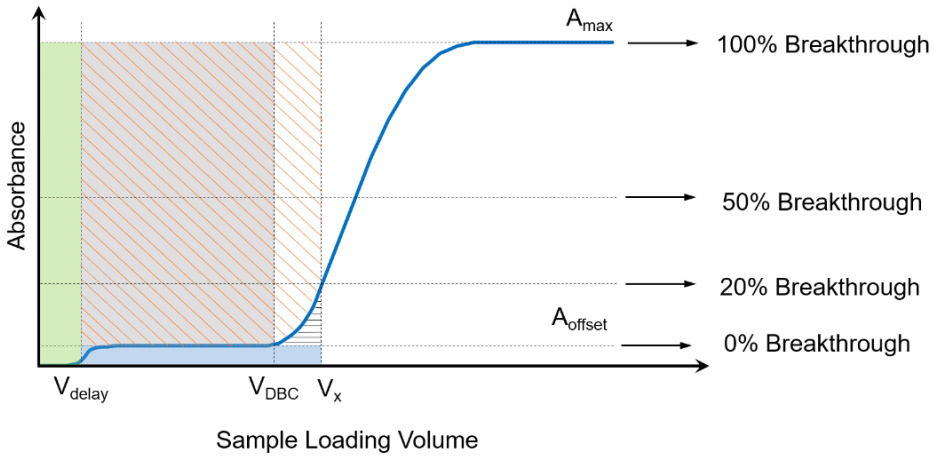

Figure 5 Typical graphic obtained for a DBC calculation. The blue curve is the UV signal. Green area is caused by the delay volume (V_{delay}), the blue area represents the amount of non-binding proteins and the grey area represents the total dynamic binding capacity at 0% breakthrough. The black hatched region represents the amount of protein that is leaked at 20% breakthrough, and the orange hatched region represents the total dynamic binding capacity at 20% breakthrough (adapted from [23]).

Therefore, the static and dynamic binding capacities depend upon the properties of the protein, the medium, and the experimental conditions. As an example, binding capacities varies according to the molecular size of the biomolecules, since a matrix with a high degree of small pores will exhibit a higher binding capacity for smaller molecules. It is then important to keep in mind that all experimental conditions (pH, ionic strength, flow rate, temperature) should be considered when comparing binding capacities of different media.

1.4. Inteins

Inteins are self-cleaving elements with the ability to excise themselves from target proteins, being this event termed protein splicing. Inteins were first described in 1990 when two research teams had begun making a significant study to insert a segment into the VMA1 gene. This gene encodes the membrane ATPase of *Saccharomyces cerevisiae*, containing approximately 1071 amino acids with 118 kDa molecular weight. However, the real molecular weight of the VMA1 protein in further tests using electrophoresis was measured to be only 67 kDa. Therefore, intein for internal and extein for the external protein sequence of the precursor protein were defined, analogously for intron and exon of mRNA, respectively. The upstream and downstream exteins were also named N-extein and C-extein, respectively [24].

Protein splicing is a post-translational processing event in which an internal protein segment, the intein, can catalyse its own excision from a precursor protein and concomitantly induce the ligation of the two flanking sequences, the exteins, to produce a mature protein [25, 26]. More than 600 inteins have been found in all three domains of life: bacteria, archaea, and eukarya [27].

Crystallography and computer modelling studies suggest that inteins consist of two different structural domains and, therefore, these protein-splicing elements are generally considered bifunctional proteins, for showing both splicing and homing endonuclease activity, essential for mobility of the intein genes and providing the means for inteins to be maintained [25, 28]. It is also described that typically the intein motifs involved in endonuclease activity are found in the central region of the intein, whereas intein motifs involved in protein splicing are found in the terminal motifs. Furthermore, it has been proposed that these functions are independent, as mutations eliminating one activity do not eliminate the other [25, 28], supporting the hypothesis that, like group I introns, mobile inteins arose by an endonuclease gene invading a sequence encoding a small functional splicing element (mini-intein) [28]. Because endonuclease activity is not required for protein splicing, mini-inteins showing splicing activity can be generated by deletion of this central domain (Figure 6) [29].

Except for the *Porphyra purpurea* DnaB intein (150 amino acids), most of the inteins that have been described are in the 400 to 500 amino acids range [25, 28].

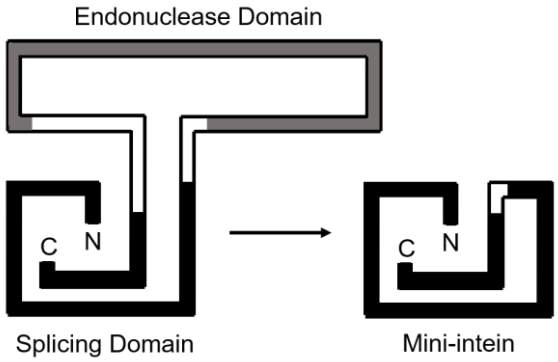

Figure 6 General intein structure with an endonuclease domain (grey area), a splicing domain (black area) and a spacer region (white area). A mini-intein showing splicing activity can be generated by deletion of the central endonuclease domain (adapted from [28]).

1.4.1. Intein Splicing Mechanism

The following Figure 7 schematize the protein splicing of an intein with an N-terminal cysteine. Step 1 is an N-S acyl shift that occurs between the C-terminal amino acid of the N-extein (coloured green) and the N-terminal cysteine of the intein (coloured red) and results in the formation of a reactive thioester. This thioester is the focus of a transesterification (step 2) reaction that results in a branched intermediate at the N-terminal cysteine of the C-extein (coloured blue). The branched intermediate is resolved by the cyclization of an asparagine residue at the C-terminus of the intein to form a succinimide group (step 3). The intein is released from the precursor and the extein residues are attached to a thioester bond. A spontaneous S-N acyl shift generates a peptide bond between the extein sequences and results in a fusion product with a peptide bond at the site of condensation (step 4) [30].

The analysis of the intein splicing mechanism, together with mechanistic studies, allowed the determination of the roles of highly conserved residues near the intein/extein junctions in the splicing reaction. These residues include the initial cysteine, serine, or threonine of the intein, that serves as a nucleophile for the splicing with a N-S acyl shift; the conserved cysteine, serine, or threonine immediately following the intein, which ligates the exteins through nucleophilic attack (transesterification); and the conserved C-terminal histidine and asparagine of the intein, which release the intein from the ligated exteins through succinimide formation (asparagine cyclization) [27, 28, 29].

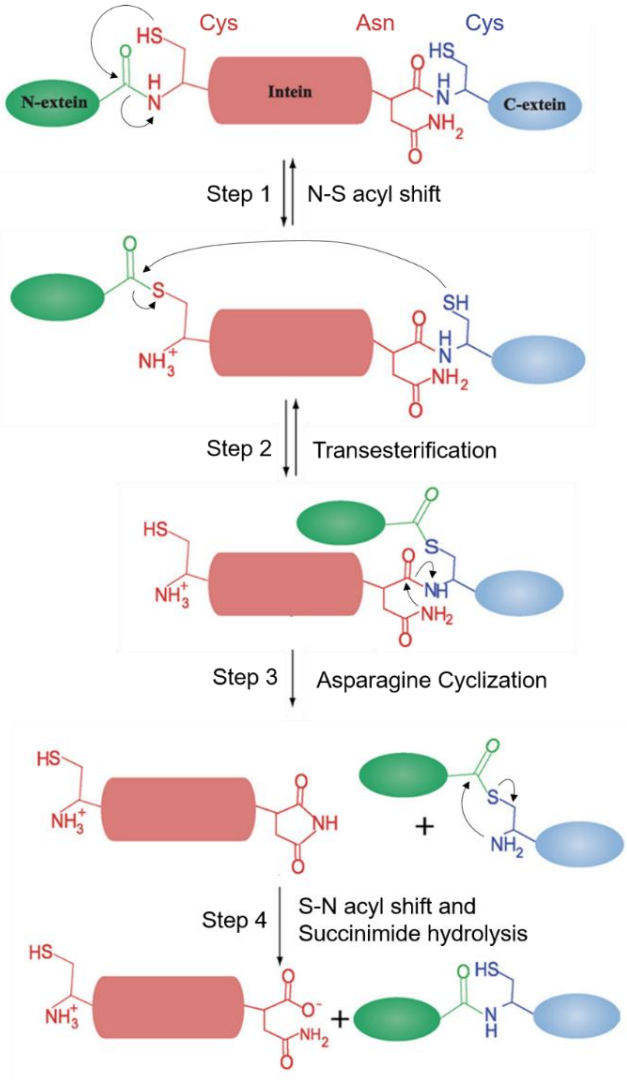

Figure 7 The four steps of protein splicing mechanism. The conserved residues are indicated: a cysteine (Cys) in the intein N-terminus, an asparagine (Asn) in the intein C-terminus and a cysteine (Cys) in the C-extein N-terminus. The N-extein, the intein and the C-extein are colored green, red and blue, respectively (adapted from [27, 30]).

However, there are still some questions remaining, namely, how are these multistep reactions coordinated in protein splicing in order to ensure orderly progression of the splicing reaction and to minimize the side reactions. Several general hypotheses exist in the literature, such as coordination by reaction rates, by mechanistic linkage, and by sequential conformational changes [31]. As an example, in a study of *Mxe* GyrA intein, Frutos *et al.* [32] have shown that transesterification and the formation of the branched intermediate (step 2) speed up C-terminal cleavage (step 3) by inducing a conformational

change in the C-terminal splicing junction, thereby linking step 2 and 3 by both reaction rate and conformational changes. However, how step 1 (N-S acyl shift) and step 2 (transesterification) are catalytically linked in protein splicing is still unclear. In fact, without a mechanistic linkage, the reactive linear thioester formed by the N-S acyl shift during step 1 could readily undergo premature N-terminal cleavage and compromise the efficiency of protein splicing [33].

1.4.2. Applying Inteins to Protein Purification

In most cases, protein purification methods with inteins do not make use of the whole intein splicing mechanism, but instead aim to induce an isolated C or N-terminal cleavage (single-splice-junction cleavage), represented in Figure 8. In fact, these inteins can be engineered through amino acid substitution, making it possible to modulate each step in the splicing pathway and convert the splicing reaction into an efficient polypeptide cleavage reaction at either splice junction [26, 34]. The challenge in using inteins for protein engineering is in learning how to control these reactions, since it may be extremely difficult to choose residues and mutations to generate specific properties [29, 35]. For this reason, a combination of rational protein design and random selection is necessary to acquire the desired characteristics for a proposed intein application [29].

As already mentioned in section 1.4.1, mechanistic studies have determined the roles of highly conserved residues near the intein/extein junctions in the splicing reaction. Therefore, mutation of these residues can be used to alter intein activity to yield isolated cleavage at one or both intein/extein junctions [29]. Also, in order for the intein to be useful for protein purification, the cleavage activity of the intein has to be attenuated *in vivo* but remain inducible *in vitro* [36].

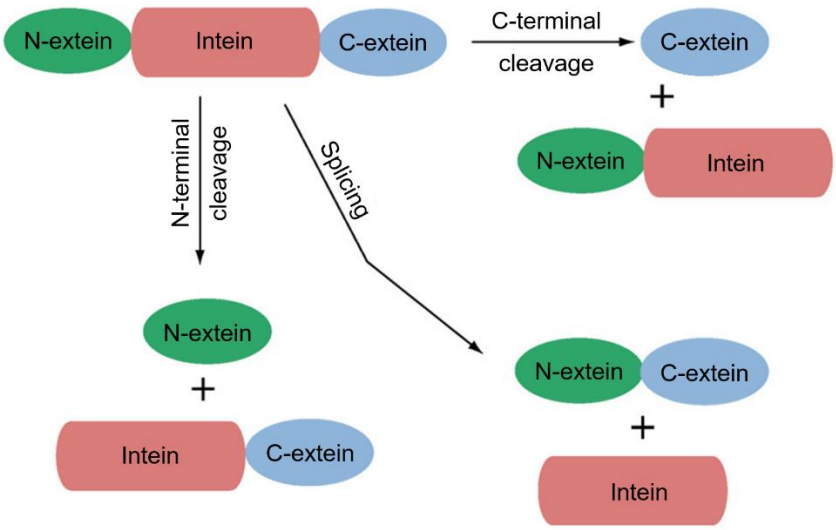

Figure 8 Side reactions of intein splicing. Inteins can be engineered through amino acid substitution to yield isolated cleavage at N or C-terminus (adapted from [27]).

1.4.2.1. The Problem

A widely used technology for recombinant protein expression and purification is to express a target protein as a fusion to an affinity tag which allows purification on an affinity column. A variety of affinity

tags have been used, including *Schistosoma japonicum* glutathione S-transferase (GST), *Escherichia coli* maltose-binding protein (MBP), *Staphylococcus* protein A, polyhistidine, calmodulin-binding peptide, antibody epitope tags, etc. [26, 35, 36]. However, several limitations exist when using this technology.

In fact, following protein purification, the affinity tag is typically cleaved from the target protein by treatment with a site-specific protease, since the tag may adversely affect the solubility or the activity of the target protein [34]. This extra proteolysis step raises a few potential limitations of this technology, making the process complex and costly [26, 35, 37]. Firstly, cleavage by proteases is not always specific and may result in cleavage at secondary sites within the target protein. Secondly, the cleavage is sometimes inefficient due to the inaccessibility of the cleavage site on the fusion protein. Thirdly, the elevated temperatures required for many proteolytic cleavage reactions may unfavorably affect protein stability or activity. And finally, an additional purification step is required to separate the target protein from the affinity tag and the protease [26].

1.4.2.2. The Solution

Protein splicing elements, named inteins, capable of catalyzing specific reactions, offer a unique alternative to cleave a peptide bond without the use of a separate protease and avoiding its associated disadvantages [34, 36].

As previously mentioned, for performing protein purification using inteins it is not desirable to use the whole intein splicing mechanism, but instead the aim is to induce an isolated C or N-terminal cleavage. Several inducible peptide bond cleavage reactions were already studied including, as an example, the cleavage reaction at the N-terminus of a modified intein from *Saccharomyces cerevisae*, the *Sce* VMA intein. This intein was modified with a single amino acid substitution, asparagine (Asn) to alanine (Ala), at the C-terminal splice junction of the intein (N454A), which showed to block splicing and C-terminal cleavage, but not the N-terminal cleavage mediated by the N-S acyl shift at cysteine (Cys) on the intein N-terminus [26]. In this study performed by Chong *et al.* [26], the target protein was fused to the N-terminus of the intein whose C-terminus was linked to a chitin-binding domain (CBD) from *Bacillus circulans* as an affinity tag. The concept was based on the observation that the modified *Sce* VMA intein could be induced to undergo a self-cleavage reaction at its N-terminal peptide linkage by thioesters, such as 1,4-dithiothreitol (DTT), β-mercaptoethanol (β-ME) or free cysteine at low temperatures and over a broad pH range. An innovating protein purification system was then developed which enabled purification of free recombinant proteins in a single chromatographic step. This was achieved cloning in-frame a target protein together with an intein and a CBD tag (step 1, Figure 9), and further purifying the stable fusion protein by adsorption onto a chitin column (step 2, Figure 9). When inside the column, the immobilized fusion protein was induced to undergo self-cleavage under mild conditions (step 3, Figure 9), resulting in the release of the target protein, while the intein-CBD fusion remained bound to the column (step 4, Figure 9). This purification method taking advantage of an intein cleavage is represented in the following Figure 9.

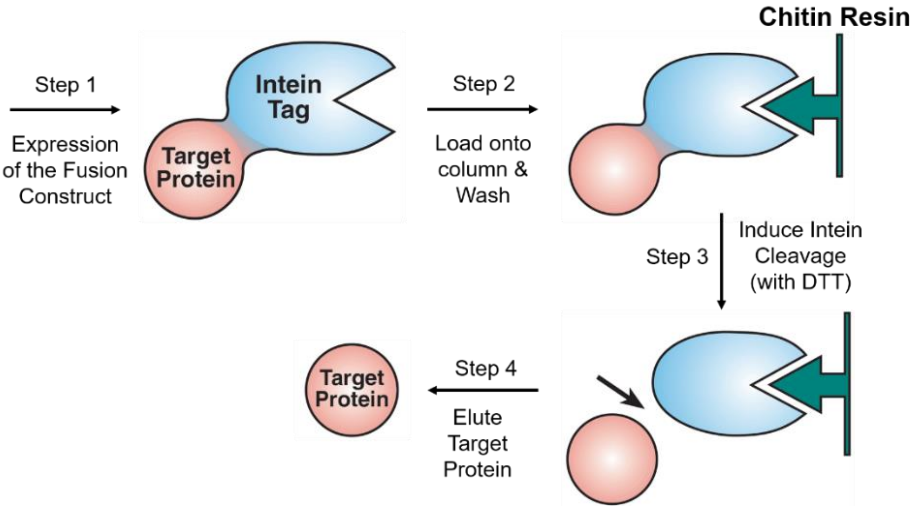

Figure 9 Schematic illustration of the intein-mediated protein purification concept with a self-cleavable affinity tag. A target protein (colored red) is fused to the N-terminus of a modified protein splicing element (intein) which is fused to an affinity tag (CBD), being this fusion construct expressed in a proper host (step 1). In this figure, the intein attached to the CBD tag is represented by the colored blue Intein Tag. The crude cell extract with the fusion is loaded onto a chitin affinity column, and the unbound proteins are washed off (step 2). The intein is then induced to undergo a self-cleavage by a chemical reagent, in the present case DTT (step 3). Lastly, the target protein is specifically released from the column and eluted as a pure protein (step 4) (adapted from [26, 38]).

Another alternative for protein purification using inteins is based on self-aggregating peptides. Elastin-like polypeptides (ELPs), for instance, are a useful tool in biotechnology and, fused with the target protein, tend to form aggregates. After inducing the precipitation of the tag, a simple solid-liquid separation method, like centrifugation, is enough to isolate the fusion proteins, whose inteins can be subsequently cleaved in a controlled manner to release the target protein into solution. After intein cleavage, the cleaved tag can be induced to precipitate, and the purified target protein collected in the supernatant after centrifugation. This procedure eliminates the chromatographic step need in most protein purification methods [34].

Taking into consideration the previous examples, it is possible to conclude a few advantages in performing intein-based purifications: no exogenous proteases are required for proteolytic removal of the tag, cleavage is specific to the intein-target protein junction and, for chromatography purifications, the fusion intein-tag remains bound to the resin after cleavage, so no further steps are required to purify the target protein [35].

1.5. Modelling in Protein Engineering

The 3D structure of enzymes can be determined essentially by X-ray crystallography and NMR spectroscopy and, for that reason, it heavily depends on computational power to process the data and require visual inspection of the results on high-end graphical computers. Graphical computers systems paved the way for the establishment of important knowledge for protein engineering, for instance, with the establishment of the structure-function relationship of proteins. Until then it was only possible to study enzymatic catalysis events by interpretation of the kinetic data under different conditions. Nevertheless, with the availability of structural data and advanced graphics computers for visualization,

it became possible to rationalize the observation at atomic detail by combining experimental and structural data [3].

An excellent example for this is the structural explanation for the reduction of the catalytic activity of subtilisin due to oxidation. Bott *et al.* [39] determined the structure of the oxidized form of subtilisin (EC 3.4.21.62) from *Bacillus amyloliquefaciens* and compared it to the nonoxidized native form. At that moment, it was possible to realize that in the oxidized form, the oxygen atom of the oxidized methionine 222 projects into the oxyanion hole and, for that reason, it is not possible for the substrate carbonyl atom to bind in an identical manner as in the native enzyme. This may not only affect the relative binding energy of the substrate, but also the catalytic rate, since the oxyanion hole has been shown to be critical for catalysis.

Besides structural knowledge, the generation of mutants is a major aspect of protein-engineering experiments. In the past, most of the mutational experiments were performed by cassette mutagenesis, a technique in which all possible natural amino acids were substituted at a given site. This technique allowed the generation of site-directed mutations, which generated mutants having, for instance, an increased substrate specificity. However, protein modelling, taking advantage of advanced graphical computer systems, may also have a huge role in predicting *in silico* the effect of a specific mutation on an enzyme performance profile and substrate specificity [3].

2. THEORETICAL BACKGROUND AND OBJECTIVES

Downstream processing of proteins is of extreme relevance when dealing with enzymes, since background impurities may have negative effects on their final application. For that reason, a range of DSM products must be purified, mostly by anion exchange chromatography which, despite being a well-known and well-established practice for protein purification, may not be very specific and has to be optimized for every product. Although purity demands for food enzymes and proteins are not very high, regulations may become stricter in the future. Nevertheless, it is important to note that there are already some specific applications in the food industry requiring high purities in the final product.

The goal of this project is to address an alternative to the conventional ion exchange chromatography which could not only yield higher purities, but could also establish a more standardized way to purify target proteins (also referred as proteins of interest). This alternative is based on the purification of target enzymes fused to a unique tag (composed by an intein fused to either an affinity tag or a precipitating tag). Making use of inteins cleaving mechanism, this intein-tag can be self-removed in the end of the process. The following Figure 10 summarizes the proposed experimental procedure to be performed in this project.

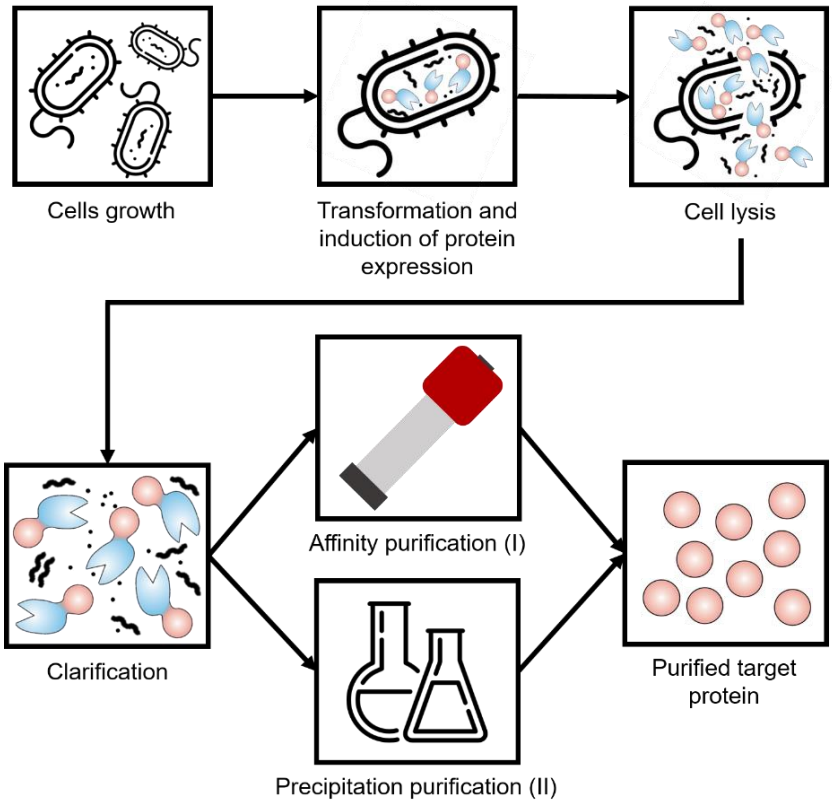

Figure 10 Regarding the fusion proteins, in red it is represented the protein of interest (POI) and in blue the intein-tag (intein fused to either an affinity tag or a precipitating tag). *E. coli* cells will be grown and further transformed with a plasmid containing the DNA sequence of the fused protein to be purified. After induction of protein expression, the cells will be lysed and submitted to a clarification step to remove cell debris. This clarified extract can then be purified either by affinity chromatography (when the fusion protein contains an affinity tag) or by precipitation (when the fusion protein contains a precipitation tag). During purification, the goal is to self-cleave the intein-tag and obtain a purified POI.

As described in the previous Figure 10, two purification methods are planned to be executed: an affinity chromatography purification (for fusion proteins expressing an affinity tag), and a precipitation purification (for fusion proteins expressing a precipitating tag). The experimental procedure on these two purification methods is schematized in Figure 11.

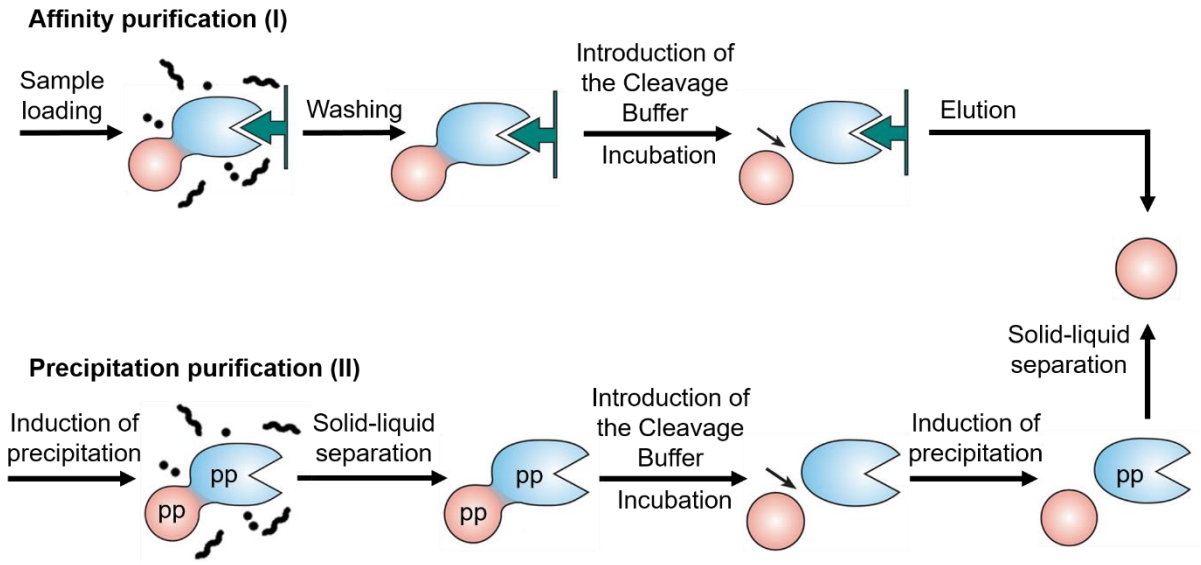

Figure 11 In red it is represented the protein of interest (POI), in blue the intein-tag (intein fused to either an affinity tag or a precipitating tag) and in green the affinity resin. For affinity chromatography purification (I) the clarified cell extract, with the POI (colored red) fused to an intein-affinity tag (colored blue), will be initially loaded onto the column, followed by a washing step to remove unbound and nonspecific bound proteins. The column will then be flushed with a buffer (Cleavage Buffer) designed to induce the cleavage of the intein-tag, followed by an incubation step. After inducing the on-column cleavage, the protein of interest can then be eluted from the column, while the intein-tag remains attached to the affinity resin. For precipitation purification (II) the clarified cell extract, with the POI (colored red) fused to an intein-precipitating tag (colored blue), will be submitted to conditions that induce the precipitation of its tag. Since this tag is fused to the POI, the whole fusion will precipitate (precipitated molecules are indicated with pp). A simple solid-liquid separation method is enough to isolate the precipitated fusion proteins which, after redissolution, can be subsequently cleaved when submitted to an incubation with Cleavage Buffer. The next step will be to induce a second precipitation and, at this moment, since the intein-tag is already cleaved from the POI, only the former will precipitate, making it possible to collect the purified POI in the supernatant after a solid-liquid separation.

As it is represented in Figure 12, six different fusion constructs (CTC01 to CTC06) will be studied. From these six different constructs, four of them (CTC01, CTC02, CTC03 and CTC04) express Enzyme α as the protein of interest (POI), attached to an intein and a purification tag, while two (CTC05 and CTC06) express Enzyme β as the POI, attached to an intein and a purification tag. In these different fusion constructs, two different inteins attached to the POI are expressed: the intein *Mxe* GyrA and the intein *Mtu* RecA. Lastly, three different purification tags are also expressed: a chitin-binding domain (CBD) and a polyhistidine tag (his-tag) for purification using affinity chromatography, and a synthetic peptide based on the natural repeat-in-toxin domain (RTX-tag) for purification after inducing a reversible precipitation. All the details regarding the used inteins and tags are described below.

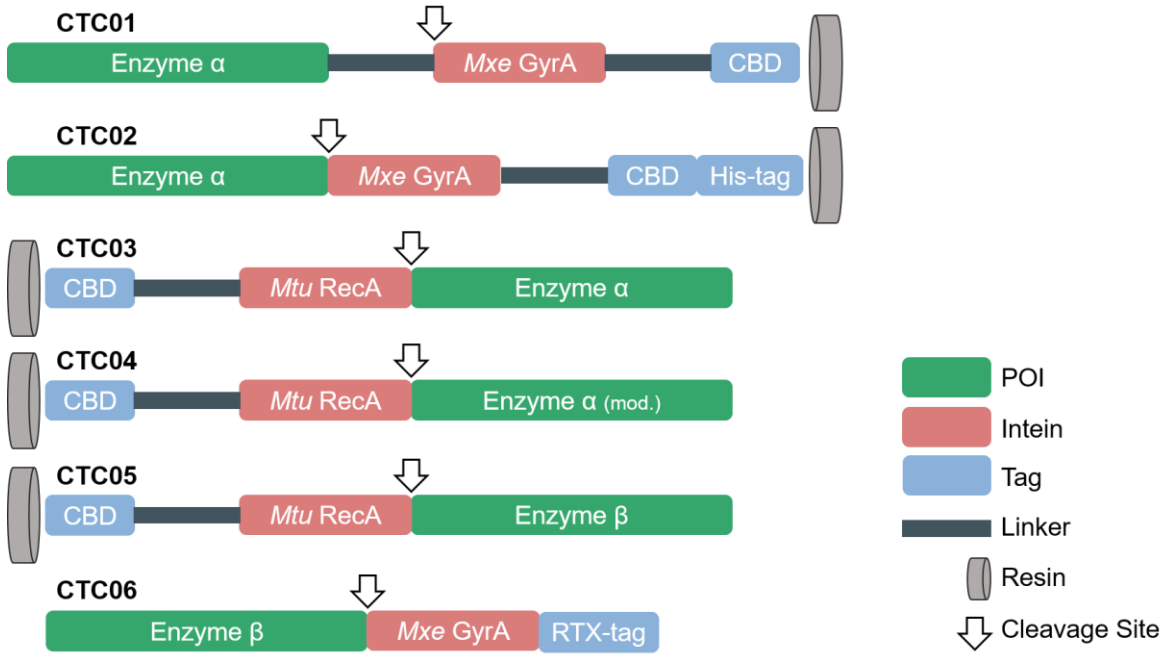

Figure 12 Scheme of the six different fusion constructs (CTC01 to CTC06) with their respective protein of interest (POI), intein, tag, linker, cleavage site and binding site to the resin. Note that CTC06 does not have a binding site to a resin since it was design designed to undergo an off-column purification. CTC04 differs from CTC03 only in a single amino acid modification (cysteine to serine) on Enzyme α N-terminus, specifically designed to favor a correct cleavage between Enzyme α and the *Mtu* RecA intein.

2.1. Proteins of Interest (POI)

2.1.1. Enzyme α

Enzyme α, typically produced by a yeast host, is a protein belonging to hydrolase class and a tetramer of four identical polypeptide chains. However, there is still some discussion around this topic, since this protein might also be active as a dimer. Each monomer has a molecular weight of around 118 kDa. The enzyme is produced intracellularly and shows an optimal activity between pH 6.5 and 7.3 and from 35 to 40°C.

2.1.2. Enzyme β

Enzyme β, typically produced by a bacterial host, is a monomeric protein belonging to hydrolase class, with a molecular weight of around 75 kDa. The enzyme is produced extracellularly and shows an optimal activity between pH 5 and 5.5 and from 60 to 70°C.

2.2. Inteins

2.2.1. *Mycobacterium xenopi* Gyrase A (*Mxe* GyrA)

The *Mycobacterium xenopi* Gyrase A (*Mxe* GyrA) intein is a natural mini-intein (198 amino acid residues), which lacks the central intein endonuclease domain and which, as a bacterial protein, can be

very highly expressed in *E. coli*. It is also an engineered intein that yields a controllable N-terminal single-splice-junction autocleavage [35].

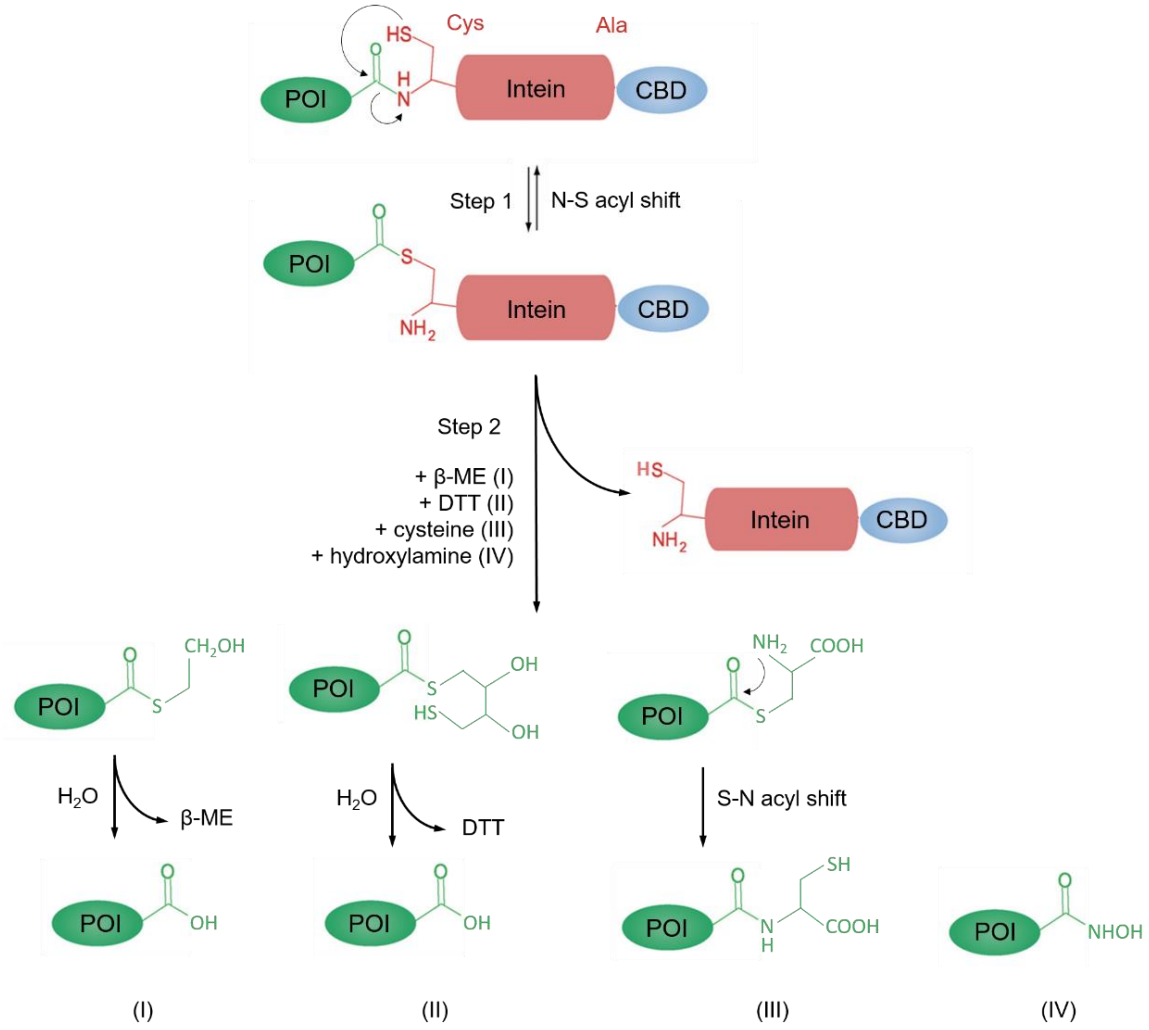

Figure 13 Chemical mechanism of the intein *Mxe* GyrA N-terminal cleavage reaction. The intein was modified by an asparagine (Asn) to alanine (Ala) mutation in its C-terminus which inactivates both splicing and C-terminal cleavage activities. This modified intein, when inserted between a protein of interest (POI) and a chitin-binding domain (CBD), catalyzes an N-S acyl shift at its N-terminal cysteine (Cys), which results in a thioester bond between the POI and the intein N-terminal cysteine (step 1). Normally, the equilibrium of the N-S shift strongly favors peptide bond formation and the fusion precursor is quite stable. The presence of β-ME (I), DTT (II), cysteine (III) or hydroxylamine (IV) shifts the N-S equilibrium by cleaving the thioester bond (step 2). After cleavage, β-ME (I) or DTT (II) is hydrolyzed from the C-terminus of the POI, while cysteine (III) or hydroxylamine (IV) remain covalently attached to the POI. In the case of cysteine (III), a subsequent S-N shift results in formation of a native peptide bond (adapted from [26]).

The natural intein splicing reaction is initiated at the N-terminal residue of the intein with a N-S acyl-shift (in inteins with an N-terminal cysteine) and the release of the intein from the precursor is mediated by the cyclization of a C-terminal asparagine residue. Mutation of the intein C-terminal reactive asparagine (Asn) to alanine (Ala) blocks splicing by preventing C-terminal cleavage. However, this mutant can still undergo the initial N-S shift, forming a thioester linkage between the N-extein (in this case the POI) and the intein (step 1, Figure 13). The precursor is stalled as a thioester intermediate whose thioester bond at the N-terminal splice junction (this bond being more reactive than a normal peptide bond) can be easily cleaved by sulfur or nitrogen nucleophiles that react with thioesters. Hence, thiols such as β-ME,

DTT or cysteine, and hydroxylamine were shown to effectively shift the N-S equilibrium by attacking the thioester, thereby inducing the N-terminal cleavage of the *Mxe* GyrA intein (step 2, Figure 13) [26, 30, 35].

2.2.2. *Mycobacterium tuberculosis* RecA (*Mtu* RecA)

The *Mycobacterium tuberculosis* RecA (*Mtu* RecA) intein has 440 amino acids, contains a conventional endonuclease domain and is expressed in *E. coli*. A series of mini-intein derivatives was tested to screen for splicing activity *in vivo* and *in vitro* and, in the end, several mini-inteins deleted for the entire endonuclease domain were shown to be capable of protein splicing in both contexts. In fact, studies showed that all information required for splicing function is carried within the first 94 and final 35 amino acids of the 440-residue intein. However, it was also shown that some deletions of the *Mtu* RecA intein did not retain activity, despite having all residues required for splicing function [28].

The cleaving mechanism for *Mtu* RecA is catalyzed by active site residues in conserved blocks of intein sequences. In fact, as previously described in section 1.4.1, inteins show some conserved residues in their structure: besides the initial cysteine (C) or serine (S) at position 1 of the intein (block-A), the histidine (H) and the asparagine (N) at the intein C-terminus (block-G) and the first residue of the C-extein, generally a cysteine (C), serine (S) or a threonine (T) (also in block-G), it is also relevant to mention two more important blocks with conserved residues in *Mtu* RecA intein structure. Block-B contains the highly conserved TxxH motif which catalyzes the N-S acyl shift, and block-F has both a conserved aspartate (D) and a conserved histidine (H) [33]. A scheme on these conserved blocks is shown in Figure 14.

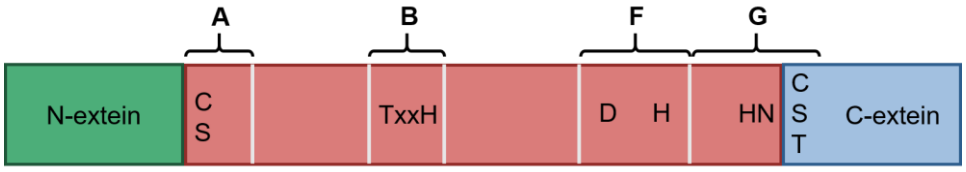

Figure 14 Conserved blocks in *Mtu* RecA intein sequence. Letters above boxes define the blocks, and letters inside indicate conserved residues. The N-extein, the intein and the C-extein are represented by the green, red and blue boxes, respectively (adapted from [33]).

The strategy to design a proper intein was based on the combination of protein engineering with random mutagenesis in order to isolate small mutant inteins with desirable splicing or cleaving properties suitable for application in affinity separations. For mutagenesis and selection studies, a mini-intein (ΔI) comprising the first 110 and the last 58 amino acids of the 440 amino acid *Mtu* RecA intein was firstly chosen. A pool of mini-inteins was generated by mutagenic PCR and further selected, where it was possible to obtain, among many others mini-inteins, the ΔI-CM showing three mutations: a conservative replacement of valine (Val) with leucine (Leu) (V67L), as well as two aspartate (Asp) to glycine (Gly) mutations, D24G and D422G (residues numbered relative to full-length *Mtu* RecA intein) [29].

The block-F aspartate D422 in the *Mtu* RecA intein has been proposed to play a pivotal role in coordinating protein splicing. The D422G mutation showed to dramatically enhance C-terminal cleavage

reaction with minimal protein splicing, indicating not only the disruption of the splicing coordination mechanism by this mutation, but also that this substitution is responsible for the elevated cleavage activity of the ΔI-CM intein. In crystal structures, D422 is positioned adjacent to both the N and C-terminal splicing junctions and has multiple conformations in various engineered *Mtu* RecA inteins. Furthermore, Belfort and Nayak groups provided a later evidence for an essential role of D422 in the formation of the branched intermediate by activating the nucleophile activity of the first residue of the C-extein [33]. The V67L mutation was found to restore the lost stability of ΔI in ΔI-CM, being this hypothesis supported by the fact that the intein is unstable in ΔI constructs and is stabilized in the ΔI-CM mutants *in vivo*. Lastly, revertant analysis of individual mutations revealed that whereas V67L restored intein stability, D24G is of no phenotypic consequence [29].

A characteristic shared by the generated mini-inteins was a strong pH sensitivity, in which cleavage rates increased as the pH was reduced. The strongest pH activation was exhibited by the ΔI-CM mutant, which also displayed a rapid C-terminal cleavage. That way, this system generated a pH-sensitive mutant intein, which obviates the need for thiol reagents, having an advantageous size and stability characteristics [29].

Therefore, the *Mtu* RecA intein used in this project correspond to the ΔI-CM mini-intein with an added cysteine (Cys) to alanine (Ala) mutation in the first residue of the intein (C1A) to enhance the suppression of N-terminal cleavage and extein ligation [29, 40].

2.3. Affinity and Precipitation Tags

2.3.1. Chitin-Binding Domain (CBD)

Affinity resins are generally expensive and, consequently, are not useful for large-scale purification. Nevertheless, tags, such as the chitin-biding domain (CBD), use as purification matrices polysaccharides that are abundantly present in nature, which could allow the development of low-cost systems for recombinant protein purification [41].

The used CBD belongs to chitinase A1 (ChiA1) from a Gram-positive bacterium, *Bacillus circulans* WL-12. Chitinase (EC 3.2.1.14) is a glycosyl hydrolase that catalyzes the hydrolytic degradation of chitin, a fibrous insoluble polysaccharide. It is possible to find chitinases in a wide variety of organisms that either possess chitin as a constituent (fungi, insects, and crustaceans), or that do not possess chitin (bacteria, plants, vertebrates), being the roles of chitinases in these organisms diverse. *Bacillus circulans* WL-12 has been reported to secrete multiple chitinases into culture medium containing chitin as an inducer. Among these chitinases, ChiA1 is thought to be the key enzyme in the chitinase system of this bacterium, since it not only is the most abundantly produced, but also exhibits the highest activity on the hydrolysis of colloidal chitin and shows a high affinity to insoluble chitin. Besides a central domain, ChiA1 contains an N-terminal catalytic domain (417 amino acid residues) and a C-terminal chitin-binding domain (45 amino acids residues), the latter required for ChiA1 to bind specifically to insoluble chitin and to hydrolyze it efficiently [42].

The present project makes use of the CBD from *Bacillus circulans*, as an affinity tag consisting of 51 amino acids. This tag is commonly available in combination with self-splicing inteins [43].

2.3.2. Polyhistidine Tag (His-tag)

To perform an immobilized metal-affinity chromatography (IMAC) a short affinity tag consisting of six polyhistidine residues were used. For the present project, histidine residues were chosen since it is the amino acid that exhibits the strongest interaction with immobilized metal ion matrices, as the electron donor groups of the histidine imidazole ring readily form coordination bonds with the immobilized transition metal. Peptides containing sequences of consecutive histidine residues are efficiently retained on IMAC column matrices and can be easily eluted by adjusting the pH of the buffer or adding either free imidazole or EDTA to the column buffer [19, 44]. Evidence suggests that some his-tag proteins that retain activity *in vivo* are inactive *in vitro* after elution with these common elution methods. Hence, a less harsh elution with L-histidine can also be performed [44].

A compromise in the length of the tag was made, since longer polyhistidine tags may result in an increased purity due to the ability to use more stringent washing steps but, on the other hand, it is also advisable to use the smallest number of histidine residues required for efficient purification to minimize possible perturbation of protein function. For these reasons, a six-histidine tag was chosen since it is reported to be generally long enough to yield high-affinity interactions with the matrix [19].

2.3.3. RTX Tag

The use of precipitations tags for non-chromatographic purification is of significant interest since chromatography is typically the most expensive step in protein purification. Thermoresponsive elastin-like peptides (ELPs) are well known and effective for many purification applications, but the necessary heating of samples or the alternative use of high salt concentrations to induce precipitation can be problematic in many situations. The calcium-responsive repeat-in-toxin (RTX) domain is found in proteins secreted through the bacterial type 1 secretion system and consists of repeats of the consensus amino acid sequence GGXGXDXUX, where X is variable, and U is a hydrophobic amino acid. One of the most well characterized RTX domains is the block V RTX domain from the adenylate cyclase toxin (CyA) of *Bordetella pertussis*, which is intrinsically disordered in the absence of calcium and forms a β roll structure in the presence of calcium [45].

The frequency of amino acids at each position of the nine amino acid repeat unit from a set of RTX-containing proteins was identified, leading to the identification of the consensus sequence GGAGNDTLY. The number of repeats of the consensus unit was also studied through a library of consensus RTX constructs, consisting of 5, 9, 13 and 17 repeats of the consensus unit fused to a target protein. The synthetic RTX domain with 17 repeats showed the best result after instantaneously precipitating upon calcium addition (optimal $CaCl_2$ concentration of 75 mM) at room temperature, and forming a pellet easy to clarify. It was also shown that pellets can be broken up upon washing, but do not redissolve until exposed to an equivalent concentration of EGTA [45].

Therefore, the RTX-tag used in CTC06 of the present project consists of 17 repeats of the consensus unit GGAGNDTLY.

3. MATERIALS AND METHODS

3.1. Materials

3.1.1. Reagents

The reagents used in the methods further described are summarized in Table 3, along with the manufacturer and minimum purity. Current laboratory material was used for preparation, handling and storage.

Table 3 Reagents used during the project, along with respective manufacturers and minimum purities.

Reagent		Manufacturer	Minimum purity
Bicine	$C_6H_{13}NO_4$	Merck	99%
Bis-Tris	$C_8H_{19}NO_5$	Sigma-Aldrich	98%
Calcium Chloride Dihydrate	$CaCl_2 \cdot 2H_2O$	Merck	99%
DNase I	EC 3.1.21.1	Merck	-
Dithiothreitol (DTT)	$C_4H_{10}O_2S_2$	Merck	99%
Histidine	$C_6H_9N_3O_2$	Sigma-Aldrich	99%
Hydrochloric Acid 4M solution	HCl	In-house preparation	-
Lysozyme	EC 3.2.1.17	Sigma-Aldrich	90% protein, 40,000 units/mg protein
Magnesium Chloride Hexahydrate	$MgCl_2 \cdot 6H_2O$	Merck	99%
Magnesium Sulfate Heptahydrate	$MgSO_4 \cdot 7H_2O$	Merck	99.5%
Potassium Chloride	KCl	Merck	99.5%
Sodium Chloride	NaCl	Merck	99.5%
Sodium Hydroxyde 4M solution	NaOH	In-house preparation	-
Tris(hydroxymethyl) aminomethane	$H_2NC(CH_2OH)_3$	Merck	99.8%
Triton™ X-100	$C_{14}H_{22}O(C_2H_4O)_{n(=9-10)}$	Sigma-Aldrich	-

3.1.2. Raw Materials

3.1.2.1. Reference Samples

Two different Reference Samples were analyzed in the present work. They consist of a fermentation broth of *E. coli* (strain A) expressing a native Enzyme α and a fermentation broth of *E. coli* (strain B) expressing a native Enzyme β. These samples were initially used in order to obtain reference values regarding the expression and stability of both proteins of interest (POI) in an *E. coli* host.

3.1.2.2. Intein Samples

For purification experiments, six different fermentation broths of *E. coli* (strain B) expressing the proteins of interest (Enzyme α and Enzyme β) attached to specific tags were used. From this point forward, this samples will be generally referred as Intein Samples. The six different fusion proteins (CTC01 to CTC06) expressed on Intein Samples were already described in the previous chapter “2.Theoretical Background and Objectives”, being a scheme represented in Figure 12.

3.1.2.3. Control Samples

As control samples, there were used three different purified enzyme formulations developed by DSM: two control samples for Enzyme α and one control sample for Enzyme β. Important to note, however, that these formulations were not obtained from an *E. coli* host and, for that reason, revealed a different protein background when compared to Reference and Intein Samples.

3.1.3. Columns and Resins

The purification of the two proteins of interest (POI) was performed based on two different affinity chromatography resins, depending on the available tag of the construct: a chitin resin for proteins expressing a chitin-binding domain (CBD), and a TALON® Superflow™ for proteins expressing a polyhistidine tag (his-tag).

3.1.3.1. Tricorn™ Column with Chitin Resin

For the purification of proteins expressing a chitin-binding domain, it was used a Tricorn™ Column (GE Healthcare, United States), made of biocompatible materials (borosilicate glass, polyetheretherketone, ethylene propylene diene monomer and perfluoro-rubber) known to exhibit excellent chemical resistance. It was used a Tricorn™ 5/50 with a 5 mm internal diameter and a bed height that ranges from 35 to 59 mm (volume from 0.69 to 1.16 mL) with one adaptor, and from 8 to 56 mm (volume from 0.16 to 1.11 mL) with two adaptors. This column also shows a maximum operating pressure of 10 MPa, 100 bar, 1450 psi and an operating temperature from 4 to 40°C [46].

The Tricorn™ column was packed with a Chitin Resin (New England Biolabs, United States), an affinity matrix used for the isolation of target proteins fused to an intein-chitin binding domain (CBD). This resin is comprised of polymerized β-1,4-N-Acetylglucosamine, with spherical chitin beads ranging in size from 20 to 100 µm. For the binding capacity, the manufacturer indicates that 2 mg of maltose-binding-protein

per mL of bed volume are released from the resin after cleavage of the maltose-binding-CBD-fusion protein. This resin may be regenerated up to 5 times and, before use, was stored at 4°C [26, 38].

3.1.3.2. HiTrap™ TALON® crude with TALON Superflow™

HiTrap™ TALON® crude prepacked with TALON Superflow™ (GE Healthcare, United States) was used for the purification of histidine-tagged recombinant proteins. This is a 1 mL volume (0.7 x 2.5 cm) commercial column made of biocompatible polypropylene, intended to not interact with biomolecules. The TALON® Superflow™ medium (GE Healthcare, United States) consists of highly cross-linked agarose beads (6%) with an immobilized chelating group. The TALON ligand is a tetra-dentate chelator charged with cobalt. This resin has a particle size distribution of 60 to 160 µm, a binding capacity up to 20 mg histidine-tagged protein per mL of medium, and a long term (one week) pH stability from pH 3 to 12, being stable in all commonly used buffers, denaturants and detergents. This prepacked column was stored at 4°C in 20% ethanol [47].

3.1.4. Buffers and Solutions

Several buffers were prepared using different chemicals in order to be specifically applied in different experiments. Biological buffers were chosen, based on the Biological Buffers Selection Guide [48], considering a suitable pK_a, usually between 6 and 8 since this is the region of most biological reactions, water solubility and stability (they should resist enzymatic and nonenzymatic degradation under the conditions of use and they should not resemble to enzyme substrates) [49]. After dissolving all components in Milli-Q water, the pH was adjusted using hydrochloric acid 4 M or sodium hydroxide 4 M solutions and measured using a PHM220 lab pH meter (Radiometer, Denmark).

3.1.4.1. General Buffers

Table 4 Composition, concentration and pH of the buffers used to stabilize Enzyme α and Enzyme β along the present project.

Buffer	Composition	Concentration	pH
A1	▪ Tris	50 mM	8.00
	▪ Potassium Chloride	100 mM	
	▪ Magnesium Chloride Hexahydrate	1 mM	
	▪ Triton X-100	0.2% (w/v)	
A2	▪ Tris	200 mM	8.00
	▪ Potassium Chloride	100 mM	
	▪ Magnesium Chloride Hexahydrate	1 mM	
	▪ Triton X-100	0.2% (w/v)	
B1	▪ Tris	50 mM	8.00
	▪ Sodium Chloride	100 mM	
	▪ Calcium Chloride Dihydrate	2 mM	
	▪ Triton X-100	0.2% (w/v)	

The three buffers described in Table 4 were specifically designed to stabilize Enzyme α (Buffer A1 and A2) and Enzyme β (Buffer B1) during the several experiments performed along this project.

3.1.4.2. Affinity Chromatography Buffers and Solutions

For the purification of the different fusion constructs (CTC01, CTC02 and CTC03), and depending on the corresponding intein, different buffers were designed with a view to either stabilize the fusions (Column Buffer) or to induce their cleavage (Cleavage Buffer). The following Table 5 summarizes the most important details of the buffers used in the Chitin Affinity Chromatography. All these buffers were prepared based on the recommendations of Instruction Manual of IMPACT™ Kit [38] and Coolbaugh and Wood studies [37].

Table 5 Composition, concentration and pH of the buffers used on the chitin affinity chromatography purifications. Depending on the purified construct a different Cleavage Buffer, specifically designed to cleave the corresponding intein was used.

Intein	***Mxe* GyrA (CTC01, CTC02)**		***Mtu* RecA (CTC03)**	
	Column Buffer pH=8.00	**Cleavage Buffer pH=8.00**	**Column Buffer pH=8.00**	**Cleavage Buffer pH=6.50**
Column and Cleavage Buffers	▪ Tris (50 mM) ▪ Potassium Chloride (100 mM) ▪ Magnesium Chloride Hexahydrate (1 mM) ▪ Triton X-100 (0.2% w/v)	▪ Tris (50 mM) ▪ Potassium Chloride (100 mM) ▪ Magnesium Chloride Hexahydrate (1 mM) ▪ Triton X-100 (0.2% w/v) ▪ DTT (50 mM)	▪ Tris (50 mM) ▪ Potassium Chloride (100 mM) ▪ Magnesium Chloride Hexahydrate (1 mM) ▪ Triton X-100 (0.2% w/v)	▪ Bis-Tris (50 mM) ▪ Potassium Chloride (100 mM) ▪ Magnesium Chloride Hexahydrate (1 mM) ▪ Triton X-100 (0.2% w/v)
Washing Buffer pH=8.00	▪ Tris (50 mM) ▪ Potassium Chloride (100 mM) ▪ Magnesium Chloride Hexahydrate (1 mM) ▪ Triton X-100 (0.2% w/v) ▪ Sodium Chloride (500 mM)			
Stripping Buffer	▪ Tris (50 mM) ▪ Potassium Chloride (100 mM) ▪ Magnesium Chloride Hexahydrate (1 mM) ▪ Triton X-100 (0.2% w/v) ▪ Sodium Hydroxide (300 mM)			
Stripping Solution	▪ Sodium Hydroxide (300 mM)			

One of the constructs (CTC02), besides having a CBD also has a his-tag. According to the instructions for HiTrap™ TALON® crude [47], Tris coordinates weakly with metal ions, causing a decrease in

capacity of this type of resins. For that reason, a different Column Buffer and Cleavage Buffer, described in the following Table 6, were prepared for the his-tag purification.

Table 6 Composition, concentration and pH of the buffers used on the immobilized metal affinity chromatography (IMAC) performed for purification of CTC02.

Intein	***Mxe* GyrA (CTC02)**	
	Column Buffer pH=8.00	**Cleavage Buffer pH=8.00**
Column and Cleavage Buffers	▪ Bicine (80 mM) ▪ Potassium Chloride (100 mM) ▪ Magnesium Chloride Hexahydrate (1 mM) ▪ Triton X-100 (0.2% w/v) ▪ Sodium Chloride (300 mM)	▪ Bicine (80 mM) ▪ Potassium Chloride (100 mM) ▪ Magnesium Chloride Hexahydrate (1 mM) ▪ Triton X-100 (0.2% w/v) ▪ Sodium Chloride (300 mM) ▪ DTT (50 mM)
Stripping Buffer pH=8.00	▪ Bicine (80 mM) ▪ Potassium Chloride (100 mM) ▪ Magnesium Chloride Hexahydrate (1 mM) ▪ Triton X-100 (0.2% w/v) ▪ Sodium Chloride (300 mM) ▪ Histidine (250 mM)	

3.2. Methods

3.2.1. Experimental Methods

3.2.1.1. Reference Samples Production

The followed method for Reference Samples production was based on a DSM standard protocol.

Cell Culture (Pre-Culture and Main-Culture)

To obtain the Reference Samples used for the initial experiments of the present project, *E. coli* (strain A) cells expressing Enzyme α and *E. coli* (strain B) cells expressing Enzyme β were inoculated in 100 mL of proper media with neomycin, in two 500 mL shake flasks. These two shake flasks (one for Enzyme α and the other for Enzyme β) were incubated overnight at 30°C and 250 rpm. The pre-cultures were then reinoculated (1:100) to two 500 mL non-baffled shake flasks, each containing 100 mL of a different proper media with neomycin. The shake flasks were incubated at 30°C and 250 rpm for 24h until induction of protein expression.

Induction of Protein Expression and Harvesting

After the cell culture step, protein expression was induced using a final concentration of 0.002% of L-arabinose, followed by an incubation for 48h at 250 rpm and 30°C. In the end, each 100 mL fermentation

broth was split into four 25 mL greiner tubes and centrifuged for 10 minutes at 5000 g. The cell pellets were stored at -20°C. Before running the subsequent experiments, each cell pellet was resuspended in 25 mL of the proper general buffer (section 3.1.4.1) - buffer A1 for cells expressing Enzyme α and buffer B1 for cells expressing Enzyme β.

3.2.1.2. Intein Samples Production

The followed method for Intein Samples production was based on a DSM standard protocol.

Gene Synthesis and Transformation

To obtain the Intein Samples, based on the desired amino acid sequence for each fusion construct (CTC01 to CTC06), each corresponding DNA sequence was initially synthetized in ATUM (DNA2.0, California) and inserted into proper plasmids developed by DSM. Afterwards, *E. coli* (strain B) cells were inoculated in a 500 mL shake flask containing 100 mL of a proper media and incubated at 30°C until an optical density measured at 600 nm (OD_{600nm}) of around 0.5 was achieved. Each plasmid containing the specific gene sequence of each construct was then transformed into competent *E. coli* cells and, after 1h recovering at 30°C, spreading the cells in agar plates with neomycin, and incubating them overnight at 30°C, six different transformants were obtained. The plates were stored at 4°C until the next cell culture step.

Cell Culture (Pre-Culture and Main-Culture)

After gene synthesis and transformation steps, a single colony from each *E. coli* variant (expressing CTC01 to CTC06) was inoculated in a 24-well plate containing 3 mL of proper media with neomycin. Reference samples (native Enzyme α and Enzyme β) were also included in this experiment to be used as controls. After the 24-well plate was sealed with a breathable seal and a lid, it was incubated overnight at 30°C, 550 rpm and 80% humidity. These pre-cultures were then reinoculated in triplicates, in two different 24-well plates containing 3 mL of a different proper media with neomycin. After sealing the plates with a breathable seal and a lid, they were incubated for 24h at 30°C until the next step of protein expression induction.

Induction of Protein Expression and Harvesting

After the cell culture step, protein expression was induced using a final concentration of 0.002% of L-arabinose, and two different temperatures (30°C and 20°C) were tested. The plate whose incubation after induction was performed at 20°C, was firstly cooled for 20 minutes in the refrigerator before L-arabinose was added. After the addition of L-arabinose, both plates were incubated for 48h at 550 rpm and 80% humidity, at 20°C (plate 1) and 30°C (plate 2).

In the end, the cells were collected by centrifugation for 10 minutes at 2750 rpm in a cooled plate centrifuge, and the cell pellets stored at -20°C. Before running the following experiments, each cell pellet was resuspended in 3 mL of the proper general buffer (section 3.1.4.1) - buffer A1 for cells expressing Enzyme α and buffer B1 for cells expressing Enzyme β.

3.2.1.3. Fusion Construct 3 (CTC03) Sample Production

At a certain stage of the project, it was decided to produce more material of the fusion construct number 3 (CTC03). For that reason, a similar procedure to Intein Samples Production (section 3.2.1.2) was performed for CTC03, with the difference that, due to the need of larger volumes, the cell culture step was performed in shake flasks instead of being performed in 24-well plates. Once again, the followed method for Fusion Construct 3 Sample production is based on a DSM standard protocol.

Cell Culture (Pre-Culture and Main-Culture)

To obtain a larger volume of CTC03 material, *E. coli* cells (strain B) expressing the fusion construct CTC03 were inoculated in a 100 mL shake flask containing 25 mL of proper media with neomycin, and further incubated overnight at 30°C and 250 rpm. This pre-culture was then reinoculated (1:100) into four 500 mL shake flasks, each containing 100 mL of a different proper media with neomycin. The four shake flaks were incubated at 30°C and 250 rpm for 24h until induction of protein expression.

Induction of Protein Expression and Harvesting

After the cell culture step, protein expression was induced using a final concentration of 0.002% of L-arabinose, followed by an incubation for 48h at 250 rpm and 20°C. In the end, the final 400 mL fermentation broth was split into eight 50 mL greiner tubes and centrifuged for 10 minutes at 5000 g. The cell pellets were stored at -20°C. Before running the following experiments, each cell pellet was resuspended in 12.5 mL (four times concentration when comparing to Reference and Intein Samples) of the general buffer A2 (section 3.1.4.1).

3.2.1.4. Cell Lysis

Sonication

The sonication method for cell lysis performed in this project was based on a standard DSM protocol specific for the available equipment. Sonication was performed by pipetting 1.5 mL of the sample to be lysed in a 2 mL plastic sample cup (2 mL Eppendorf Tube®) using an ultrasonic disintegrator Soniprep 150 (MSE, England) adjusted to 10 cycles, 10 seconds on and 5 seconds off. It was set a power output which kept an amplitude value of 10 microns. During sonication, the sample cup was kept in melting ice.

Lysozyme Treatment

The lysozyme treatment for cell lysis performed in this project was based on a standard DSM protocol. For this lysis method, a 50 mM Tris-HCl lysis buffer, pH 7.5, supplemented with 2 mg/mL of lysozyme, 0.1 mg/mL DNase I and 25 µM of $MgSO_4 \cdot 7H_2O$ was firstly prepared.

Lysozyme treatment was achieved by adding 800 µL of lysis buffer to a cell pellet obtained after centrifugation of a 1 mL of fermentation broth. For the specific experiment performed in this project, a pellet corresponding to 7 mL fermentation broth was brought in suspension in 5.6 mL lysis buffer and the mixture was then incubated for 45 minutes at 37°C.

3.2.1.5. Chitin Affinity Chromatography

The purification experiments using Inteins Samples CTC01, CTC02 and CTC03 were performed following slightly different protocols, so that different conditions could be tested during the initial screening. All following protocols (including packing, different phases, buffers, solutions, incubation times, used column volumes and flowrates) were established based on the Instruction Manual of IMPACT™ Kit [38] and on the studies performed by Coolbaugh and Wood [37] and Chong *et al*. [36].

Packing the Chitin Resin into a Tricorn™ Column

The chitin beads (New England Biolabs, United States) were stored at 4°C in a slurry of 20 % ethanol. The bottle containing this chitin resin was well shook to resuspend the beads, and then around 1.5 times of the desired column bed volume of slurry was slowly added to the Tricorn™ Column 5/50 (GE Healthcare, United States), in order to avoid the formation of air bubbles in the packed volume. For all the chitin purification experiments performed in this project, the bed volume after packing was 1 mL. Note that references to column volume (CV), from this point forward, refer to the bed volume. After guaranteeing that the column was leveled, the excess slurry volume was allowed to drain by gravity for around 30 minutes, the column was closed with the top adjustable cap and further connected to ÄKTA avant 25 system from GE Healthcare, under the control of UNICORN 7.3 software (GE Healthcare, United States). For the final packing, the column was subjected to a continuous flow of Column Buffer starting with 0.02 mL/min for 5 CV, which was then stepwise increased by 0.02 mL/min until 0.12 mL/min for 20 CV.

Preparation of the Loading Sample

After production of the samples to be purified (sections 3.2.1.2 and 3.2.1.3) and respective lysis by sonication (section 3.2.1.4), the obtained lysates were clarified by centrifugation for 20 minutes at 15000 g and 4°C, using a centrifuge 5430 R (Eppendorf, Germany).

Affinity Chromatography – CBD v1

This method was used for purification of the Inteins Samples CTC01 and CTC03 after respective production and preparation. The column was initially equilibrated at 0.08 mL/min with Column Buffer during 5 CV. After this first step, 1 mL of sample was then loaded into the top of the column via the sample loop at 0.02 mL/min, emptying the sample loop with 4 mL of Column Buffer, and collecting, from this moment forward, the unbound material in the flow through in 0.5 mL fraction volumes (0.5 CV) within 96-well plates (GE Healthcare, United States) placed on the cassettes of ÄKTA avant 25 tray.

The next washing step was performed with 6 CV of Column Buffer at 0.08 mL/min, followed by the introduction of Cleavage Buffer phase, in which the column was washed using 5 CV of Cleavage Buffer at 0.08 mL/min to evenly distribute the Cleavage Buffer throughout the column. In order to induce the on-column cleavage, the column was incubated at room temperature for different times, depending on the intein to be cleaved: for CTC01, with a *Mxe* GyrA intein, the incubation took 40 hours, while for CTC03, with a *Mtu* RecA intein, the incubation took 22.5 hours.

The elution of the cleaved target protein was performed washing the column at 0.08 mL/min with 6 CV of Column Buffer for CTC01 (with a *Mxe* GyrA intein), and with 6 CV of Cleavage Buffer for CTC03 (with a *Mtu* RecA intein). To strip the remaining material still bound to the chitin resin after elution (in theory both the uncleaved fusion precursor protein and the intein-tag), the column was washed at 0.08 mL/min during 5 CV using 0.3 M NaOH in Column Buffer (Stripping Buffer).

Lastly, after each run, the resin was regenerated at 0.08 mL/min with 3 CV of Stripping Solution (0.3 M NaOH) followed by a 30 minutes pause at room temperature to allow the resin to soak. The regeneration was then repeated with an additional 7 CV of Stripping Solution and with a subsequent wash with Milli-Q water during 7 CV. A re-equilibration step with Column Buffer was finally performed at 0.08 mL/min during 5 CV. All collected fractions were then analyzed via NuPAGE® electrophoresis and Enzyme α activity assay.

All the suggested flowrates were described by Chong *et al.* [36] for an Econo-Column® (2.5 cm x 10 cm, Bio-Rad, United States) packed with 30 mL of chitin beads (New England Biolabs, United States). When scaling-up, the flow rates (mL/min) should be increased, maintaining, on the other hand, the flow velocity (cm/h) [17]. Therefore, all the described flow rates in [36] were converted to their respective linear flow velocity, which were then reconverted into flow rates applied to the Tricorn™ Column 5/50 (GE Healthcare, United States) of 5 mm of internal diameter used in the present project.

Affinity Chromatography – CBD v2

This method was used for purification of the Inteins Samples CTC01 and CTC02 after respective production and preparation. The second version of the Chitin Affinity Chromatography method is very similar to the first one (“Affinity Chromatography - CBD v1”), with only a few changes. For that reason, in the following description, only the changes made to “Affinity Chromatography - CBD v1” will be mentioned.

The initial equilibration and sample loading were performed as already described in “Affinity Chromatography - CBD v1”. The following washing step was performed with 6 CV of Column Buffer supplemented with 0.5 M NaCl (Washing Buffer) at 0.08 mL/min, in order to reduce nonspecific binding of other *E. coli* proteins. The induction of on-column cleavage, with the introduction of the Cleavage Buffer and subsequent incubation, was also performed as previously described

Since the full capacity of the column was being poorly used, and with a view to obtaining more concentrated eluted samples in the first fractions, the elution was performed as described in “Affinity Chromatography - CBD v1”, but with a reverse flow (upflow elution).

Lastly, both the Stripping and Regeneration phases were performed as described before. All collected fractions were then analyzed via NuPAGE® electrophoresis and Enzyme α activity assay.

Breakthrough Curve for Dynamic Binding Capacity Determination

For breakthrough studies, a pure fusion sample was not available, and for that reason, the present test was performed with Fusion Construct 3 (CTC03), whose production (section 3.2.1.3) and preparation

were already described. The chitin resin manufacturer [38] indicates that 2 mg of maltose-binding-protein (MBP) per mL of bed volume are released from the resin after cleavage of the maltose-binding-CBD-fusion protein. Considering this theoretical binding capacity for MBP, an extrapolation was done to Enzyme α.

After production, lysis and preparation of the Fusion Construct 3 (CTC03) sample, an Enzyme α activity assay was performed on the initial sample to determine its Enzyme α concentration. An initial equilibration of the column (as described in “Affinity Chromatography - CBD v1”) was performed and then, based on the theoretical maximum binding capacity, 16 mL of this initial sample were applied into the column using the sample pump at 0.02 mL/min. From this moment forward, the unbound material in the flow through was collected in 0.5 mL fraction volumes (0.5 CV) within 96-well plates (GE Healthcare, United States) placed on the cassettes of ÄKTA avant 25 tray.

After the sample loading, all the further phases described in “Affinity Chromatography CBD - v1” for Enzyme α purification (washing, introduction of the Cleavage Buffer, incubation to induce on-column cleavage, elution and stripping) were performed before proceeding to the column regeneration. Due to the overloading of the column, a longer washing phase of 18 CV after sample application had to be performed until the 280 nm absorption reading returned to the baseline. All collected fractions were then analyzed via NuPAGE® electrophoresis and Enzyme α activity assay.

Affinity Chromatography – CBD v3

The following method was performed for purification of the Fusion Construct 3 (CTC03), an identical sample used in the dynamic binding capacity determination test, after respective production (section 3.2.1.3) and preparation. This third version of the Chitin Affinity Chromatography method was developed to reproduce the first one (“Affinity Chromatography - CBD v1”), but with a higher sample loading to use around 70% of the maximum binding capacity experimentally found. For that reason, in the following description, only the changes made to “Affinity Chromatography - CBD v1” will be mentioned.

After production, lysis and preparation of the Fusion Construct 3 (CTC03) sample, an Enzyme α activity assay was performed on the initial sample to determine its Enzyme α concentration. An initial equilibration of the column (as described in “Affinity Chromatography CBD - v1”) was performed and then, based on the previously determined experimental maximum binding capacity, 7.5 mL of this initial sample were applied into the column using the sample pump at 0.02 mL/min. From this moment forward, the unbound material in the flow through was collected in 0.5 mL fraction volumes (0.5 CV) within 96-well plates (GE Healthcare, United States) placed on the cassettes of ÄKTA avant 25 tray.

After sample loading, all the further phases described in “Affinity Chromatography - CBD v1” for Enzyme α purification (washing, introduction of the Cleavage Buffer, incubation to induce on-column cleavage, elution and stripping) were performed before proceeding to the column regeneration. All collected fractions were then analyzed via NuPAGE® electrophoresis and Enzyme α activity assay.

3.2.1.6. Immobilized Metal Affinity Chromatography (IMAC)

For his-tag purification, only one method was performed, and was established based on the Instructions for HiTrap™ TALON® crude 1 mL and 5 mL TALON Superflow™ [47], combined with the Instruction Manual of IMPACT™ Kit [38] and Coolbaugh and Wood studies [37].

Preparation of the Loading Sample

After production of the samples to be purified (section 3.2.1.2) and respective lysis by sonication (section 3.2.1.4), the obtained lysates were clarified by centrifugation for 20 minutes at 15000 g and 4°C, using a centrifuge 5430 R (Eppendorf, Germany).

Affinity Chromatography – His-tag

This method was used for purification of the Intein Sample CTC02 after respective production and preparation. The system flow was set to 1 mL/min for all phases. The column was initially washed with 5 CV of Milli-Q water to wash out the ethanol, and was then equilibrated with Column Buffer during 8 CV. After these first steps, 1 mL of sample was loaded into the top of the column via the sample loop, emptying it with 5 mL of Column Buffer, and collecting, from this moment forward, the unbound material in the flow through in 1 mL fraction volumes (1 CV) within 96-well plates (GE Healthcare, United States) placed on the cassettes of ÄKTA avant 25 tray.

The next washing step was performed with 15 CV of Column Buffer, followed by the introduction of Cleavage Buffer phase, in which the column was washed using 5 CV of Cleavage Buffer to evenly distribute the Cleavage Buffer throughout the column. In order to induce the on-column cleavage, and since CTC02 contains a *Mxe* GyrA intein, the column was incubated at room temperature for 40 hours.

The elution of the cleaved target protein was performed washing the column with 6 CV of Column Buffer with a reverse flow (upflow elution). To strip the remaining material still bound to the chitin resin after elution (in theory both the uncleaved fusion precursor protein and the intein-tag), the column was washed with 8 CV using 250 mM Histidine in Column Buffer (Stripping Buffer).

A re-equilibration step with Column Buffer was finally performed with 8 CV. All collected fractions were then analyzed via NuPAGE® electrophoresis and Enzyme α activity assay.

3.2.2. Analytical Methods

3.2.2.1. NuPAGE® Electrophoresis

Electrophoresis was performed using NuPAGE® Bis-Tris 4-12% (1.0 mm, 10 wells) polyacrylamide gels (Thermo Fisher Scientific, United States) designed for optimal separation and resolution of small to medium sized proteins (1.5-300 kDa) under denaturing gel electrophoresis conditions [50].

The samples were prepared by mixing, in 1.5 mL Eppendorf Tubes®, 65 µL of the sample to be analyzed (when needed, the sample was first diluted), 25 µL of a preset NuPAGE® LDS Sample Buffer (4X) and

10 µL of a preset NuPAGE® Reducing Agent (10X). After mixing, the samples were incubated for 10 minutes at 70°C and 550 rpm in a thermomixer comfort (Eppendorf, Germany).

A volume between 5 to 30 µL of prepared sample was loaded in each well and the gels were placed in an Invitrogen™ Novex Mini Cell (Thermo Fisher Scientific, United States), which was further connected to an electrophoresis PowerEase® 500 Power Supply (Thermo Fisher Scientific, United States). The upper buffer chamber of the Invitrogen™ Novex Mini Cell was filled with 200 mL of 1X NuPAGE® MES or MOPS SDS Running Buffer and 500 µL of NuPAGE® Antioxidant, while the lower buffer chamber was filled with 600 mL of 1X NuPAGE® MES or MOPS SDS Running Buffer.

The use of 4-12% Bis-Tris Gels combined with MES Running Buffer allows a clear separation of 2.5 to 200 kDa proteins, while with MOPS Running Buffer allows a clear separation of 14 kDa to 200 kDa. The used protein ladder was SeeBlue® Plus2 Prestained Standard (Thermo Fisher Scientific, United States) which contains 10 proteins markers ranging from 3 to 198 kDa.

The electrophoresis was performed at a constant voltage of 200 V, 120 mA and 25.0 W for 35 to 45 minutes. After the run, the gels were stained with a coomassie based staining solution, Instant Blue™ (Expedeon, United Kingdom), for around 1 hour, and then destained by submerging and washing the gels in water and gently mix them overnight using a diffusion-destaining-apparatus (Desaga GmbH, Germany).

Lastly, the imaging of the gels was obtained using a G:Box (Syngene, India) scanner controlled by GeneSys image capture software.

3.2.2.2. Enzyme α Activity Assay

The assay for determining the activity of Enzyme α performed in this project is based on an internal DSM protocol. This method is applicable for the determination of Enzyme α activity in fermentation or downstream processing samples. One unit of Enzyme α (α-unit) corresponds to the amount of Enzyme α that liberates 0.81 µmol of glucose per minute using the substrate under pH 6.45 and 37°C. Hence, in this method, glucose was used as a measure of Enzyme α activity.

After incubation of Enzyme α (POI) with its substrate at pH 6.45 and 37°C, the latter was converted into a few products, including glucose. After this incubation time, the reaction was then stopped by the addition of a stop reagent, followed by an incubation. Taking advantage of a diagnostic reagent kit for quantitative *in vitro* determination of glucose, Glucose Hexokinase FS kit (DiaSys, Germany), it was possible to use the released glucose as a measure for enzymatic activity. The Glucose Hexokinase FS contains 2 different reagents: R1 (Tris buffer 100 mM at pH 7.8 with 4mM Mg^{2+}, 2.1 mM ATP and 2.1 mM NAD) and R2 (4mM Mg^{2+}, ≥7.5 kU/L Hexokinase (HK) and ≥7.5 kU/L Glucose-6-phosphatedehydrogenase (G6P-DH)). The subsequent addition of these two reagents (in a ratio 4:1, R1:R2) converted the released glucose into Gluconate-6-P in two steps (Equations 1 and 2), during which NADH is also formed.

$$\text{Glucose} + \text{ATP} \xrightarrow{\text{HK}} \text{Glucose-6-phosphate} + \text{ADP} \quad (1)$$

$$\text{Glucose-6-phosphate} + \text{NAD}^+ \xrightarrow{\text{G6P-DH}} \text{Gluconate-6-P} + \text{NADH} + \text{H}^+ \quad (2)$$

After a proper incubation time, the resulting increase in absorbance at a wavelength of 340 nm was used as a measure for the released glucose ($A_{340nm\ sample}$).

However, $A_{340nm\ sample}$ needed to be corrected with a blank measurement ($A_{340nm\ blank}$), in which the sample and all the reagents were added in the same amounts, but in a different order so that the reaction between Enzyme α sample and the substrate did not proceed. For this blank measurement, also happening at pH 6.45 and 37°C, the stop reagent was initially added to the substrate so that the following reaction with Enzyme α could not occur. The subsequent addition of R1 and R2 (from Glucose Hexokinase FS) in a ratio R1:R2=4:1, followed by an incubation allowed the blank absorbance measurement at 340 nm ($A_{340nm\ blank}$).

Enzyme α activity was calculated through a glucose calibration line obtained using a glucose standard with an officially assigned content. Five dilutions of the glucose standard were prepared. The absorbance at 340 nm (A_{340nm}) for each standard dilution was measured after submitting all these five solutions to an identical test as the one performed to calculate the $A_{340nm\ blank}$ for Enzyme α samples, with the only difference that, in this case, each standard dilution (instead of Enzyme α sample) was added.

After calculating the exact amount of glucose (in µmol) that was present in the assay for the different calibration points of glucose standard, it was possible to create a calibration line by plotting the calculated amount of glucose (µmol) *versus* the corrected absorbance at 340 nm. This calibration line must be fitted applying a linear regression using

$$Y = aX + b \quad (3)$$

Where Y is the corrected A_{340nm} and X the amount of glucose in µmol. The slope, a, of this line was used to determine the activity of Enzyme α samples using Equation 4.

$$\frac{\alpha\ units}{mL\ sample} = \frac{A_{340nm\ sample} - A_{340nm\ blank}}{a} \times \frac{1000}{Vs} \times 0.81 \times \frac{Df}{t} \quad (4)$$

Where, $A_{340nm\ sample}$ is the sample absorbance at 340 nm in sample test, $A_{340nm\ blank}$ is the sample absorbance at 340 nm in blank test, a is the slope of the calibration line (abs/µmol), Vs is the volume of sample in assay (µL), 1000 is the conversion factor from µL to mL, 0.81 is the factor from unit definition, Df is the total dilution factor and t is the incubation time (min) between Enzyme α sample and the substrate.

All the dilutions of the samples and standards were performed using a diluter Hamilton Microlab 600 Series (Hamilton, United States) or Gilson verity® 4120 (Gilson, United States). The activity assay was performed using a Konelab Arena 30 clinical analyzer (Thermo Fisher Scientific, United States).

3.2.2.3. Enzyme β Activity Assay

This method is applicable for the determination of Enzyme β activity in fermentation or downstream processing samples. One unit of Enzyme β (β-unit) corresponds to the amount of Enzyme β that releases 1 µmol of p-nitrophenol (pNP) per minute using the end blocked pNP substrate at pH 5.20 and 37°C. For this reason, in this method, pNP was used as a measure of Enzyme β activity.

Enzyme β activity was determined at 37°C and pH 5.20 using an end blocked p-nitrophenyl substrate. After incubation of Enzyme β (POI) with the end blocked pNP substrate, the latter was hydrolyzed into some products, including free p-nitrophenol. The reaction was then terminated and color developed by the addition of a stop reagent followed by a proper incubation time. After this last incubation, the resulting increase in absorbance at a wavelength of 405 nm ($A_{405nm\ sample}$) due to the release of p-nitrophenol was used as a measure of enzymatic activity.

However, $A_{405nm\ sample}$ needed to be corrected with a blank measurement ($A_{405nm\ blank}$), in which the sample and the reagents were added in the same amounts, but in a different order so that the reaction between Enzyme β and the substrate did not proceed. For this blank measurement, also happening at pH 5.20 and 37°C, the stop reagent was initially added to Enzyme β sample so that the following reaction with the substrate did not occur. The blank absorbance was then measured at 405 nm ($A_{405nm\ blank}$).

Enzyme β activity was calculated from a molar extinction coefficient determination, through a calibration line obtained using a p-nitrophenol standard solution with an officially assigned content. Six dilutions of the p-nitrophenol (pNP) standard were prepared. The absorbance at 405 nm (A_{405nm}) for each standard dilution was measured after submitting all these six solutions to an identical test as the one performed to calculate the $A_{405nm\ blank}$ for Enzyme β samples, with the only difference that, in this case, each standard dilution (instead of Enzyme β) was added.

At that moment, it was possible to prepare a calibration line by plotting the absorbance at 405 nm (A_{405nm}) *versus* the corresponding calculated p-nitrophenol concentration of the standard dilutions (mM). This calibration curve must be fitted applying a linear regression using Equation 3.

Where *Y* is the A_{405nm} and *X* the concentration of pNP in the incubation solution (mM). This curve was used to determine the molar extinction coefficient (MEC) of pNP according to Equation 5.

$$\varepsilon_{pNP} = 1000 \times a \tag{5}$$

Where ε_{pNP} is the molar extinction coefficient of p-nitrophenol at a wavelength of 405 nm (abs·L·mol^{-1}·cm^{-1}), 1000 is the conversion factor from L/mmol to L/mol and a is the slope of the pNP calibration line (abs/mM). Note that Equation 5 follows Lambert-Beer Law, in which the molar extinction coefficient (ε) is a measure of how strongly a chemical specie or substance absorbs light at a particular wavelength, being an intrinsic property of the chemical substance that is dependent on their chemical composition and structure.

The molar extinction coefficient of p-nitrophenol was used to calculate the activity of Enzyme β samples using Equation 6.

$$\frac{\beta\ units}{mL\ sample} = \frac{A_{405nm\ sample} - A_{405nm\ blank}}{\varepsilon_{pNP}} \times 1000 \times \frac{Vf}{Vs} \times \frac{Df}{t} \quad (6)$$

Where $A_{405nm\ sample}$ is the sample absorbance at 405 nm in sample test, $A_{405nm\ blank}$ is the sample absorbance at 405 nm in blank test, ε_{pNP} is the molar extinction coefficient of p-nitrophenol at a wavelength of 405 nm (abs·L·mol^{-1}·cm^{-1}), 1000 is the conversion factor from mmol to µmol, Vf is the assay final volume (µL), Vs is the volume of sample in assay (µL), Df is the total dilution factor and t is the incubation time (min) between Enzyme β sample and the substrate.

All the dilutions of the samples and standards were performed using a diluter Hamilton Microlab 600 Series (Hamilton, United States) or Gilson verity® 4120 (Gilson, United States). The activity assay was performed using a Konelab Arena 30 clinical analyzer (Thermo Fisher Scientific, United States).

3.2.2.4. Mass Spectrometry

Mass Spectrometry was performed for protein identification, based on the analysis of peptides generated by proteolytic digestion. This was achieved by comparing the experimentally determined MS peak mass values with the predicted molecular mass values of the peptides generated by a theoretical digestion of each protein in a database.

Two different digestions were performed: in-gel digestion with trypsin, and in-solution digestion with both trypsin and endoproteinase Asp-N. In gel-digestion allows the analysis of highly complex samples, but with less sequence coverage and, for that reason, was only used for protein identification purposes. On the other hand, in-solution digestion, used for samples with lower complexity (purified samples), shows a higher sequence coverage and was used for identification of specific peptide sequences in a target protein.

3.2.3. Molecular Modelling

For Molecular Modelling, a Maestro software with the Bioluminate® package (Schrödinger, United States) was used. Each PDB ID corresponding to Enzyme α, intein *Mxe* GyrA, intein *Mtu* RecA and to the chitin-binding domain (CBD) was collected from RCSB Protein Data Bank [51], and further uploaded on Maestro software. To model a predicted structure for CTC01 and CTC03, several other similar structures (also extracted from RCSB Protein Data Bank) were used as a template, also taking advantage of some software tools, namely the energy minimization and the protein preparation wizard.

4. RESULTS AND DISCUSSION

4.1. Initial Experiments with Reference Samples

The present project began with the analysis of the Reference Samples in order to have a better insight on the expression, stability and optimal conditions of both Enzyme α and Enzyme β when produced in *Escherichia coli*. Although these samples were not expressing the fusions constructs to be purified, these first steps revealed a great importance, since both proteins of interest (POI) are not natively produced in this bacterial host. The microorganism *E. coli* is, in fact, commonly used for recombinant protein production, and was chosen as the host for the present project for being considered a fast screening system that shows a fast growth and high protein yields. On the other hand, its inability to easily secrete recombinant proteins into the extracellular medium is a drawback for industrial production processes [52]. Nevertheless, purification is most challenging when the POI is produced intracellularly since complete lysis is required, which ends up releasing all the intracellular proteins. For that reason, the use of such a host, allowed the study of a worst-case scenario for purification.

4.1.1. Cell Lysis

After production of the Reference Samples (section 3.2.1.1), the first step was to test cell lysis under conditions that would not only keep high activity levels of the POI but would also prevent the dissociation of the fusion constructs in the future experiments with Intein Samples. Mechanical methods are very efficient to lyse a wide range of cells, being high pressure homogenization one of the most widely used methods for large scale microbial disruption [53]. Additionally, bead milling also suggests being an accessible method for cell disruption with good potential for industrial scale-up [54]. On the other hand, non-mechanical methods, such as enzymatic digestion, are usually more specific and gentler. Enzymatic lysis with, for instance, lysozyme that reacts with the peptidoglycan layer of bacterial cells, is an alternative approach with possible uses in industry [53, 54]. However, the cost of lysozyme and its subsequent presence as a contaminant in later downstream processing steps have been impeding a wide large-scale use of this enzyme [55]. Furthermore, when using non-mechanical lysis, one should consider whether the added chemicals/reagents are compatible with the subsequent steps in the purification process, also guaranteeing that they do not lead to premature cleavage of the fusion constructs.

For the aforementioned reasons, two different lysis methods were tested. Firstly, a sonication method was chosen as a non-mechanical method for being the most representative method of large scale, at laboratory scale. Secondly, as a non-mechanical method, lysozyme treatment was also chosen. Although the chitin resin manufacturer [38] states that egg white lysozyme is not recommended for cell lysis since it is known to bind and degrade chitin, this method was still performed in order to have data to compare with sonication method.

After performing cell lysis, the samples before and after removal of cell debris were analyzed via both NuPAGE® electrophoresis and an activity assay for the POI. Although the analysis on these two extracts

do not allow to calculate an exact cell lysis yield, it gives an indication on the proportion of POI present in the soluble and insoluble fractions after performing each cell lysis method.

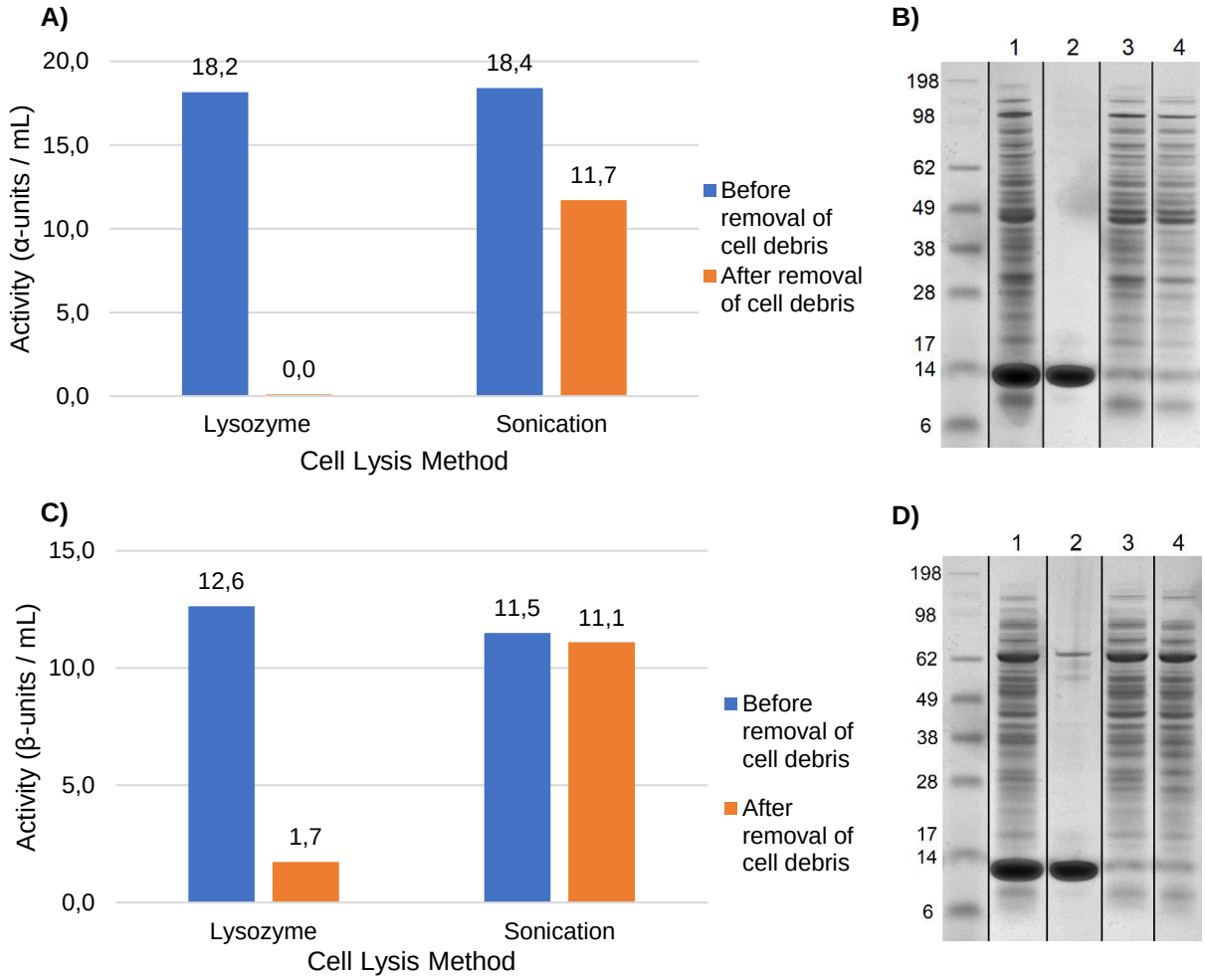

Figure 15 Cell lysis (lysozyme treatment and sonication) of the Reference Sample expressing Enzyme α and Enzyme β as the protein of interest. A) Enzyme α activity (α-units/mL) on samples after lysis and before removal of cell debris (blue) and after removal of cell debris (orange). B) Polyacrylamide gel (5 µL loaded volume) of the Enzyme α samples after lysozyme treatment before removal of cell debris (lane 1) and after removal of cell debris (lane 2), and samples after sonication before removal of cell debris (lane 3) and after removal of cell debris (lane 4). C) Enzyme β activity (β-units/mL) on samples after lysis and before removal of cell debris (blue) and after removal of cell debris (orange). D) Polyacrylamide gel (5 µL loaded volume) of the Enzyme β samples after lysozyme treatment before removal of cell debris (lane 1) and after removal of cell debris (lane 2), and samples after sonication before removal of cell debris (lane 3) and after removal of cell debris (lane 4). Samples before removal of cell debris were prepared following the NuPAGE® electrophoresis procedure but, after incubation with the LDS sample buffer and the reducing agent, were centrifuged and only their supernatant was loaded in the gel.

In order to infer about possible cell lysis during fermentation, the obtained supernatant after harvesting cells during Reference Samples production (section 3.2.1.1) was also analyzed, and neither Enzyme α nor Enzyme β activity was found. Therefore, it is possible to conclude that the observed lysis was only due to either lysozyme treatment or sonication.

The fact that in both lysis methods the lysates before removal of cell debris showed a significant activity, and knowing that, as previously mentioned, *E. coli* is not able to naturally secrete recombinant proteins without signal peptides [52, 56], one can conclude that the performed lysis methods worked to some extent. Since it was possible to measure activity in lysates before removal of cell debris, the substrate

was in contact with Enzyme α or Enzyme β intracellularly produced. However, when analyzing the lysozyme treated samples after removal of cell debris, not only the POI (tracked through activity assays) but also all the remaining proteins (visible in the gels of Figure 15 B) and D)) remained in the insoluble fraction. Some of these proteins could have remained attached to cell debris, precipitated due to aggregation, be present in inclusion bodies, or one may also assume that complete cell lysis did not occur but, instead, only a cell wall permeabilization was achieved. This last possibility could explain the penetration of the substrate into the permeabilized cell during the activity assay. Regarding sonication, the analysis of the samples after removal of cell debris showed that most of the POI remained in the soluble fraction. The higher yield observed for Enzyme β when compared to Enzyme α could be related to the fact that the latter protein is almost twice as long and forms a complex quaternary structure.

If the samples contain large amounts of nucleic acids, the target protein's mobility may be affected, or the complex may form large insoluble aggregates [57], which could explain the inefficacy of the lysozyme treatment. Although DNase was included in the lysozyme treatment to avoid this possibility, either the inefficacy of this enzyme or the non-inclusion of RNase, may also explain the observed lysis differences between the two tested lysis methods. On the other hand, nuclease treatment is not required for sonicated material since sonication breaks down nucleic acids, rendering them incapable of forming intact protein-binding domains [57, 58].

A second possible explanation for this difference is based on the fact that lysozyme reacts with the peptidoglycan layer and breaks glycosidic bonds. For that reason, gram-positive bacteria can be directly exposed to lysozyme, while the outer membrane of the gram-negative bacteria may need to be removed before exposing the peptidoglycan layer to the enzyme. Although lysozyme treatment was combined with a first freeze-thaw cycle, it might not had been enough to completely disrupt the outer membrane, and so the inclusion of detergents might also play a significant role in this treatment [53].

Overall, these first experiments contributed to establish a sonication protocol that was further used along the present project.

4.2. Fusion Constructs Screening with Intein Samples

4.2.1. Optimization of Expression

According to [38], different induction conditions should be tested if *in vivo* cleavage of the intein is suspected. For *Mxe* GyrA mini-intein, literature suggests that induction at lower temperature results in increased solubility, improved folding and subsequent thiol induced cleavage [38]. Additionally, for *Mtu* RecA mini-intein, literature [37] mentions that a critical step to prevent premature intein cleavage is to cool the culture to, at least, room temperature prior to inducing expression. The subsequent expression is then suggested to be performed from 15 to 25°C. Regarding the expression time, a compromise had to be made since yields are typical higher with longer expression times, however, there is also a higher risk for premature cleavage with longer expression. For the aforementioned reasons, two different expression temperatures were tested (20°C and 30°C) in order to study the influence of this parameter on fusion constructs' integrity and expression. After production and cell lysis, all the Intein Samples

expressing the fusion constructs (CTC01 to CTC06) were analyzed: the activity assays (Figure 16 A)) included samples before and after removal of cell debris, while NuPAGE® electrophoresis (Figure 16 B) and C)) was only performed with the supernatants after removal of cell debris.

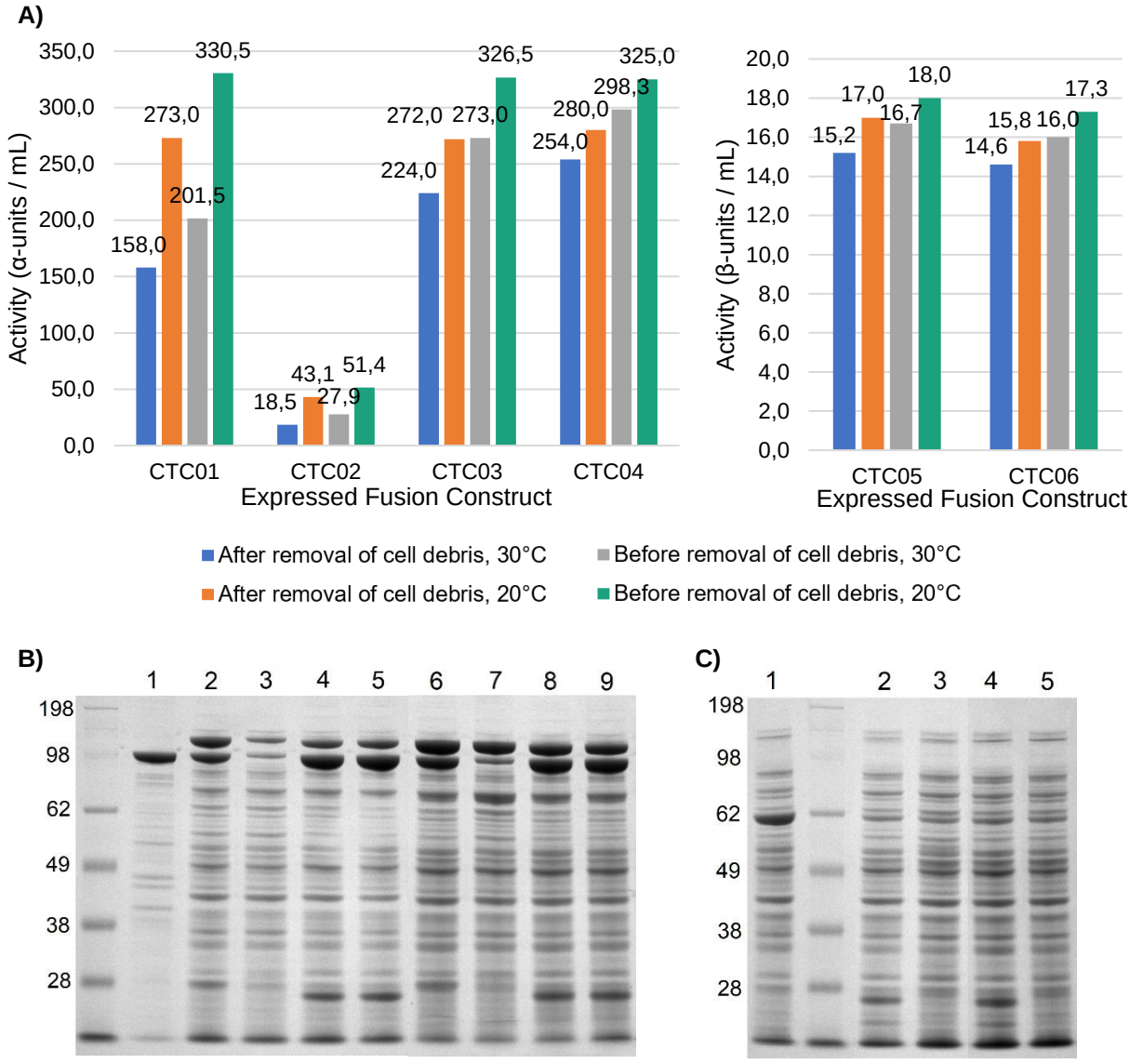

Figure 16 Initial screening of the six different fusion constructs (CTC01 to CTC06) expressed at 20°C and 30°C. A) Enzyme α / β activity on Intein Samples, induced at 30°C and at 20°C, after sonication, before removal of cell debris and after removal of cell debris. B) Polyacrylamide gel (5 µL loaded volume) of the Intein Samples expressing Enzyme α fusions (CTC01 to CTC04) after sonication and removal of cell debris. Lane 1: Enzyme α Control Sample (Enzyme α MW ≈ 100 kDa); lane 2: CTC01 induced at 30°C; lane 3: CTC02 induced at 30°C; lane 4: CTC03 induced at 30°C; lane 5: CTC04 induced at 30°C; lane 6: CTC01 induced at 20°C; lane 7: CTC02 induced at 20°C; lane 8: CTC03 induced at 20°C; lane 9: CTC04 induced at 20°C. C) Polyacrylamide gel (5 µL loaded volume) of the Intein Samples expressing Enzyme β fusions (CTC05 and CTC06) after sonication and removal of cell debris. Lane 1: Enzyme β Control Sample (Enzyme β MW ≈ 65 kDa); lane 2: CTC05 induced at 30°C; lane 3: CTC06 induced at 30°C; lane 4: CTC05 induced at 20°C; lane 5: CTC06 induced at 20°C.

Firstly, as it was already expected form the previous cell lysis study (section 4.1.1), Figure 16 A) shows that samples before removal of cell debris always revealed a higher POI activity than the corresponding samples after removal of cell debris. Note that this comparison must be done between samples induced at the same temperature, which means comparing, for each Intein Sample of Figure 16 A), the blue columns with the grey ones and the orange columns with the green ones. Secondly, Figure 16 A) also

revealed a very interesting result: Intein Samples expressed at 20°C (orange and green columns) always revealed a higher POI activity than their corresponding samples induced at 30°C (blue and grey columns). So, assuming a constant POI specific activity, these higher activities suggest higher expressions levels.

The following Table 7 summarizes the approximated molecular weights of the different fusions constructs for a better analysis of the polyacrylamide gels (Figure 16 B) and C)).

Table 7 Approximated molecular weights of the POI, intein and tag composing each fusion construct. The molecular weights of the linkers and of the his-tag were ignored, due to their insignificant impact on the total molecular weight of each fusion construct.

Fusion Construct	POI (kDa)	Intein (kDa)	Tag (kDa)	Total (kDa)
CTC01	118	21	6	145
CTC02	118	21	6	145
CTC03	118	19	6	143
CTC04	118	19	6	143
CTC05	75	19	6	100
CTC06	75	21	14	110

Figure 16 B) shows two clear bands around 100 kDa for each Intein Sample expressing an Enzyme α fusion (CTC01 to CTC04). Comparing these bands with the native Enzyme α Control Sample (Figure 16 B), lane 1, MW(Enzyme α) ≈ 100 kDa), one could assume that the lower band correspond to the native Enzyme α, and the upper band corresponds to the fusion construct (Enzyme α attached to an intein-tag). This conclusion is supported by the fact that the bands corresponding to the fusion constructs are positioned in the gel as expected, considering the molecular weights of the intein-tags of 27 kDa for CTC01 and CTC02, and of 25 kDa for CTC03 and CTC04.

However, having bands corresponding to the native Enzyme α on the Intein Samples was not initially expected, and their presence may be due to two different reasons: either native Enzyme α was individually expressed during Intein Samples production, or the expressed fusion construct underwent early cleavage. The latter seems to be the most reasonable explanation, since when analyzing Figure 16 B), it is also possible to identify, especially for CTC01, CTC03 and CTC04, intense bands between 25 to 27 kDa, whose molecular weight is in accordance with the corresponding intein-tag. These observations suggest that the intact fusions constructs were expressed but, at some point, underwent premature cleavage. In fact, early cleavage of the fusion constructs can occur since these inteins are known to be temperature sensitive and, during sample production, the higher temperatures may have an impact on their stability. Furthermore, even though samples were always kept on ice during the cell lysis procedure, sonication is also known to heat up samples. Lastly, the cleavage of the *Mtu* RecA mini-intein is induced by decreasing the pH from 8 to 6.5 and, during sample production, the internal pH of the cell might be closer to the cleaving pH, also leading to a possible premature cleavage.

The influence of the NuPAGE® electrophoresis protocol in the cleavage of the fusion constructs was also studied. In fact, this protocol includes a sample incubation step in the presence of DTT for 10

minutes at 70°C. The combination of heating with DTT, even for a short incubation time, could indeed induce a partial cleavage of the inteins and, consequently, of the fusion constructs. For that reason, all the Intein Samples were analyzed via NuPAGE® electrophoresis incubating them at 70°C without the pre-treatment with DTT, and without both the pre-treatment with DTT and the incubation at higher temperature (data not shown). Overall, the ratio between the fusion construct and the native POI did not change, which indicates that the early cleavage does not occur while performing the electrophoresis.

The higher expression of the Intein Samples induced at 20°C shown in Figure 16 A) is also supported by Figure 16 B), where is possible to compare the different Enzyme α fusion constructs (CTC01 to CTC04) induced at 30°C (lane 2 to 5) and at 20°C (lane 6 to 9). In fact, samples induced at 20°C showed not only a higher Enzyme α expression, but also a higher ratio (fusion construct):(native Enzyme α), suggesting that temperature during sample production is critical to the stability of the fusion constructs, and play a role on early cleavage.

Worth to mention is the completely unexpected behavior of the Intein Sample expressing the fusion construct 2 (CTC02), which showed a very low Enzyme α activity when compared to the other fusion constructs (Figure 16 A)). In fact, CTC02 only differs from CTC01 for having an extra his-tag and for lacking a linker between Enzyme α and *Mxe* GyrA intein. For that reason, either this linker, initially designed to promote a correct cleavage between Enzyme α and the intein, plays an important role in protein folding; or the his-tag affects the protein folding of the whole construct, consequently affecting its activity and expression.

The results for Enzyme β fusion constructs (CTC05 and CTC06) were not so clear. Comparing the Intein Samples (Figure 16 C), lanes 2 to 5) with the native Enzyme β Reference Sample (Figure 16 C), lane 1, MW(Enzyme β) ≈ 65 kDa), it is difficult to clearly identify the bands corresponding to either the fusion constructs or to the native Enzyme β. However, for CTC05 it is still possible to identify a more intense band around 25 kDa, whose molecular weight is in accordance with the corresponding intein-tag, which might suggest an early cleavage of the fusion construct. The expression of these fusion constructs (CTC05 and CTC06), either at 20°C or at 30°C, were substantially low which could suggest folding issues. This unclear expression would create difficulties in tracking the intein cleavage through electrophoresis and thus, it was preferred to continue the studies with Intein Samples expressing the Enzyme α fusion constructs.

Enzyme α fusion constructs CTC01, CTC02 and CTC03 were considered the most interesting samples to use in the following purification experiments. CTC01 had a very high expression using the *Mxe* GyrA mini-intein. Although CTC02 had a very poor expression, this construct seemed promising for combining a CBD tag with a his-tag, which could allow an interesting comparison between the affinity purification using the two different tags. Lastly, CTC03 and CTC04, both having a *Mtu* RecA mini-intein, are very similar constructs. The only difference is a single amino acid modification (cysteine to serine) on Enzyme α N-terminus of CTC04, specifically designed to favor a correct cleavage between Enzyme α and the intein. Since both constructs showed similar activity and similar premature cleavage, it was decided to move forward with CTC03 since its Enzyme α was not modified. Overall, performing protein expression

at 20°C yielded better results than at 30°C and, consequently, all the samples used in the following experiments were expressed at 20°C.

4.3. Affinity Chromatography with Intein Samples

The following affinity chromatography purification experiments were performed in order to test the feasibility of the intein cleavage mechanism of the most promising constructs (CTC01, CTC02 and CTC03) in a chromatography column. The main goal at this stage was not to completely optimize the purification system but, instead, try to understand its feasibility and choose, between CTC01, CTC02 and CTC03, the most promising fusion construct to study further.

The ÄKTA avant 25 system can measure the UV absorbance at 280 nm (A_{280nm}) in order to follow protein content, but it is not able to specifically track Enzyme α. Thus, all the collected fractions were analyzed via Enzyme α activity assays in order to specifically track this POI. Also, analysis via NuPAGE® electrophoresis was also fundamental to not only infer about the integrity of the fusion constructs since, in principle, activity assays cannot distinguish the native Enzyme α from the corresponding fusion construct, but also to have a qualitative measure on the purity of each sample.

4.3.1. Fusion Construct 1 (CTC01)

Fusion Construct 1 (CTC01) combines the chitin-binding domain (CBD) tag with the *Mxe* GyrA mini-intein to yield a self-cleaving purification tag. Enzyme α was genetically fused downstream of the tag, generating a fusion protein (CTC01) whose purification was tested using a chitin resin. In principle, intein N-terminal cleavage is induced by incubation with 50 mM DTT. The effect of the -1 residue (residue immediately preceding the splicing competent *Mxe* GyrA intein) on splicing and cleavage of the *Mxe* GyrA intein was studied in [35, 38] and, for the specific residues present in CTC01, the higher cleavage yield was obtained during a 40 hours incubation at 23°C.

After packing a 1 mL bed volume of chitin resin into a Tricorn™ Column 5/50 and preparing the Intein Sample expressing CTC01 at 20°C, the purification of the fusion CTC01 was initially performed following the method “Affinity Chromatography – CBD v1” described in section 3.2.1.5. From the sample flow through during sample application, until stripping phase all the unbound material was collected in the column flow through in 0.5 mL fraction volumes (0.5 CV) to be then analyzed via Enzyme α activity assay and NuPAGE® electrophoresis.

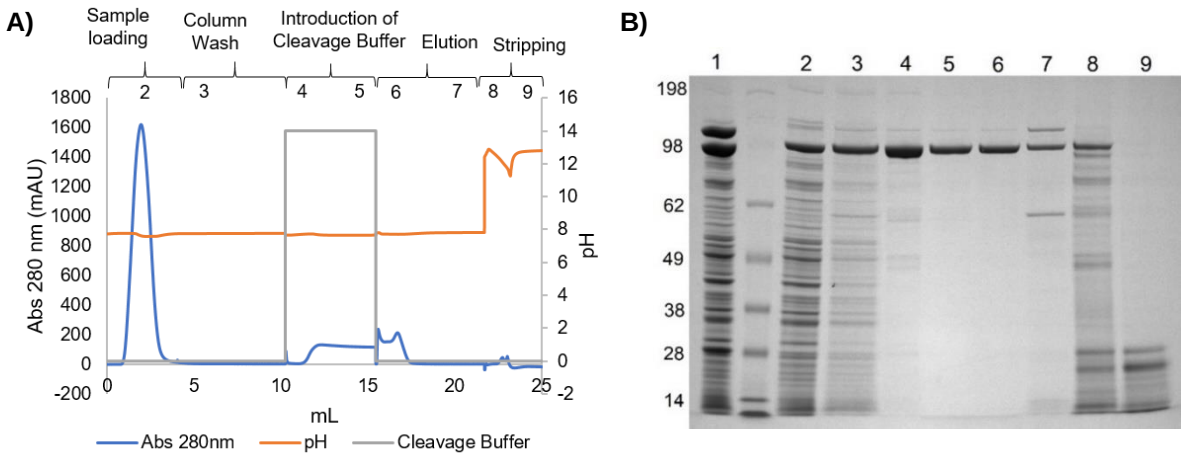

Figure 17 A) Chromatogram of CTC01 purification following "Affinity Chromatography – CBD v1". The fractions collected during column wash, introduction of the Cleavage Buffer, elution and stripping were split: Column Wash 1 corresponds to the initial 2 CV of column wash, while Column Wash 2 corresponds to the remaining 4 CV of column wash; Cleavage Buffer 1 corresponds to the initial 2 CV of introduction of the Cleavage Buffer, while Cleavage Buffer 2 corresponds to the remaining 3 CV of introduction of the Cleavage Buffer; Elution 1 corresponds to the initial 2 CV of elution, while Elution 2 corresponds to the remaining 4 CV of elution; lastly, Stripping 1 corresponds to the initial 2.5 CV of stripping, while Stripping 2 corresponds to the remaining 3 CV of stripping. The grey line represents the step (0 to 100%) introduction of the Cleavage Buffer. The numbering below each sample phase is corresponding to the gel numbering. B) Polyacrylamide gel of the initial loaded sample and of the several collected fractions during chromatography. Lane 1: Initial loaded sample; lane 2: Sample loading flow through; lane 3: Column Wash 1; lane 4: Introduction of Cleavage Buffer 1; lane 5: Introduction of Cleavage Buffer 2; lane 6: Elution 1; lane 7: Elution 2; lane 8: Stripping 1; lane 9: Stripping 2. Before running the electrophoresis, some samples' concentration was firstly adjusted and the loaded amount in each well was also not constant, with a view to achieving a proper protein concentration in each lane.

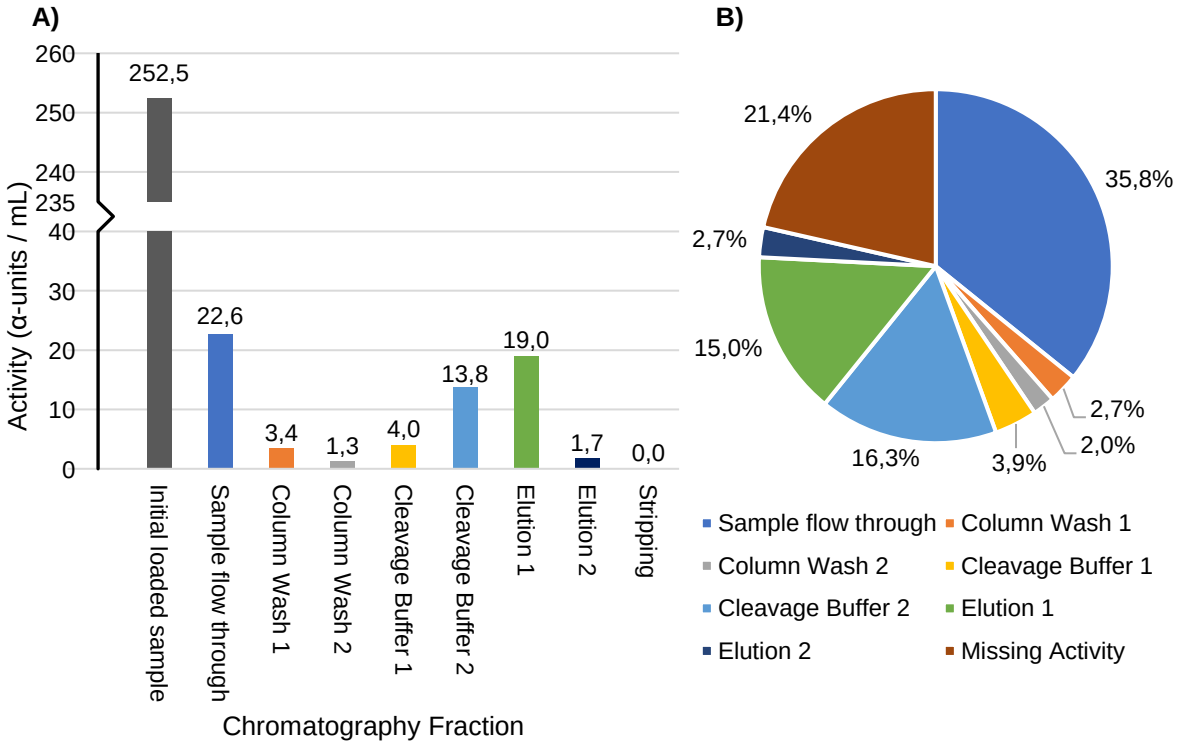

Figure 18 Enzyme α activity assay results after affinity purification of CTC01 following the "Affinity Chromatography – CBD v1" procedure. A) Activity measurements on the initial sample loaded into the column and on the whole fractions collected during chromatography. B) Percentage of Enzyme α activity present in each fraction in relation to the total amount of Enzyme α activity initially loaded into the column.

Figure 18 summarizes the Enzyme α activity results. Figure 18 A) represents Enzyme α activity (α-units/mL) in the initial sample loaded into the column, and in the whole fractions collected during chromatography. Based on these activity concentrations, and knowing the volume of each collected fraction, the total activity of Enzyme α in each fraction was calculated. Figure 18 B) represents the percentage of Enzyme α activity present in each fraction in relation to the total amount of Enzyme α activity initially loaded into the column. This way, it was possible to establish an Enzyme α activity balance to the affinity column in order to follow Enzyme α along the whole chromatographic process. In fact, assuming that there is no inactivation during the purification process and that native Enzyme α shows the same molar catalytic activity (α-unit/mmol) as the fusion construct, the activity loaded into the column must be equal to the activity that leaves the column in the several different fractions.

It is important to note that it was not possible to measure activity on the stripping phase due to the high pH that inactivates Enzyme α (Figure 18 A)). When calculating the activity balance, it was impossible to close it, since part of the activity seemed to be kept inside the column after the elution phase. Hence, this “missing activity” was calculated in order to close the activity balance and, considering the two previous assumptions, it should be related to stripping phase.

Comparing the initial sample (Figure 17 B), lane 1) with the sample loading flow through (Figure 17 B), lane 2) it is possible to observe that the upper band disappeared, which indicates that this band corresponds to the fusion construct expressing a CBD that bound to the chitin resin. On the other hand, the lower band representing the native Enzyme α, for not having a CBD was simply washed out from the column. These two bands were analyzed by mass spectrometry (MS) which confirmed their identity. The more intense band with a MW ≈ 27 kDa present in the initial sample (lane 1) was also analyzed by MS which confirmed to be the cleaved intein-tag.

Figure 17 B) shows that, as expected by the initial Fusion Constructs Screening (section 4.2), the initial loaded sample (lane 1) is partially cleaved, being possible to estimate a 1:1 (fusion construct):(native Enzyme α) ratio. Taking this into account, and assuming an identical molar catalytic activity for the CTC01 and for the native Enzyme α, one could assume that at least 50% of the activity loaded into the column (corresponding to the native Enzyme α) should be washed from the column in the first phases of chromatography. Figure 18 B) shows that 40.5% of the activity initially loaded into the column was washed out from the column during sample application and column wash, which is below the expected value, possibly indicating that some native Enzyme α might show some kind of interaction with the stationary phase. However, a significative amount of Enzyme α activity (20.2%) left the column during the introduction of the Cleavage Buffer. This result, together with the high activity found in the fractions collected during Introduction of Cleavage Buffer 2 (Figure 18 A)), suggests that the introduction of the Cleavage Buffer, even before the 40 hours incubation, may already induce some premature cleavage of the fusion construct. Note that during the introduction of the Cleavage Buffer there is also an increase in the 280 nm absorbance (Figure 17 A)) most likely due to the presence of DTT in the Cleavage Buffer.

Regarding purity, the fractions obtained during Introduction of Cleavage Buffer 2 (before the 40 hours incubation) and Elution 1 (after incubation) represented, respectively, in lanes 5 and 6 of Figure 17 B), did qualitatively show a very high purification factor. On the other hand, the last collected fractions during

elution (Elution 2 fraction, represented in lane 7 of Figure 17 B)) began to show the presence of both the uncleaved fusion construct and an unknown protein (≈ 60 kDa), additionally to the native Enzyme α.

A purification yield for a general sample Φ was calculated based on the following Equation 7.

$$\eta_{\phi}^{purification} = \frac{Activity_{\phi} \times Volume_{\phi}}{Activity_{IS} \times Volume_{IS} \times R/100} \times 100\% \quad (7)$$

Where $\eta_{\Phi}^{purification}$ is the purification yield of sample Φ, $Activity_{\phi}$ is the measured activity on sample Φ (α-units/mL), $Volume_{\phi}$ is the total volume of sample Φ (mL), $Activity_{IS}$ is the measured activity on the initial loaded sample (α-units/mL), $Volume_{IS}$ is the total volume of initial sample loaded into the column, and R represents the estimated percentage of fusion construct with respect to both fusion and native Enzyme α present in the initial sample (%). The value of R was estimated based on the (fusion construct):(native Enzyme α) ratio observed after running the initial sample in a polyacrylamide gel, and takes into account the fact that part of the measured activity in the initial loaded sample comes from the native Enzyme α that, for not having a CBD, cannot be purified.

For the present case, assuming a 1:1 ratio (corresponding to $R = 0.5$), it was possible to calculate for Elution 1 fraction a $\eta_{Elution\ 1}^{purification} = 30\%$. Assuming that CTC01 chromatography could be optimized in order to avoid the premature washing of purified native Enzyme α during the introduction of the Cleavage Buffer, then it was possible to calculate for the combination of Introduction of Cleavage Buffer 2 and Elution 1 fractions a promising $\eta_{Cleavage\ Buffer\ 2\ +\ Elution\ 1}^{purification} = 63\%$.

It is important to note that a significative amount of Enzyme α activity (21.4%) was missing in the activity balance. Assuming no inactivation during chromatography and an identical molar catalytic activity (α-unit/mmol) for the cleaved Enzyme α and for CTC01, this missing activity should correspond to the retained activity inside the column after elution phase, which might indicate that complete intein cleavage was not achieved. However, this conclusion seems a little contradictory with the previously hypothesized premature cleavage during the introduction of the Cleavage Buffer. This suggests some heterogeneity in the cleavage tendency due to, for instance, any modification of the cysteine that could likely hamper cleavage.

The high pH achieved during stripping phase did not allow to measure activity in this fraction (Figure 18 A)). However, after inducing the intein on-column cleavage and elute the cleaved Enzyme α, it is expected that some uncleaved fusion precursor and the cleaved intein-tag remain bound to the column. Although the stripping lanes of Figure 17 B) show the washing of the cleaved intein-tag (≈ 27 kDa), no fusion construct was washed from the column. Instead, native Enzyme α was washed out which might indicates that either all fusion construct molecules were cleaved but, for some reason, did not leave the column; or not all fusion construct molecules were cleaved, but under a very high pH (achieved during stripping phase) its cleavage was also induced.

4.3.1.1. Influence of Reverse Flow Elution and Stringent Washing Conditions

The chitin resin manufacturer [38] indicates that 2 mg of maltose-binding-protein (MBP) per mL of bed volume is released from the resin after cleavage of the maltose-binding-CBD-fusion protein. Considering this theoretical binding capacity for MBP and extrapolating it to Enzyme α (assuming the binding of the same number of molecules, ignoring possible steric hindrance effects due to the higher dimensions of Enzyme α compared to MBP and hypothesizing that, in average, one tag per Enzyme α dimer binds to the resin) it was possible to estimate that around 11 mg of Enzyme α per mL of bed volume should be released from the resin after cleavage of the Enzyme α fusion constructs. Thus, assuming a 100% yield on the intein cleavage, it is possible to estimate a theoretical maximum binding capacity of 11 mg of Enzyme α per mL of bed volume. The measured activity content of CTC01 initial sample (α-units/mL) was converted to Enzyme α concentration (mg/mL) using Enzyme α specific activity (α-units/mg), showing that only 3 to 6% of the theoretical column capacity was being used. For that reason, "Affinity Chromatography CBD v2" was performed with a view to understanding the influence of reverse flow elution (up-flow elution) that, in theory, would improve peak shapes and allow to obtain more concentrated purified fractions [20]. Moreover, a washing step with higher salt concentration after sample loading, was also included to reduce nonspecific interactions with the chitin resin [19, 20].

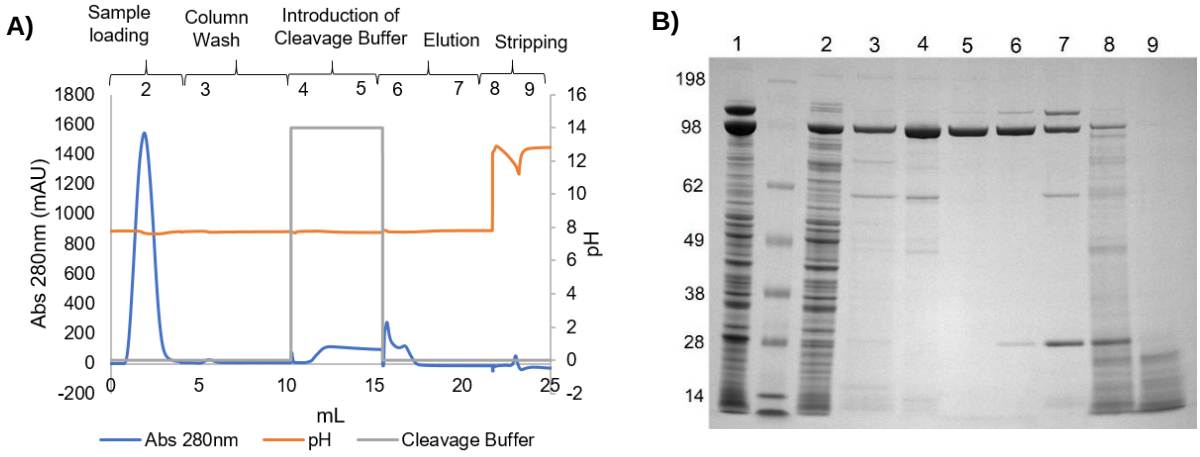

Figure 19 A) Chromatogram of CTC01 purification following "Affinity Chromatography – CBD v2". The fractions collected during column wash, introduction of the Cleavage Buffer, elution and stripping were split: Column Wash 1 corresponds to the initial 2 CV of column wash, while Column Wash 2 corresponds to the remaining 4 CV of column wash; Cleavage Buffer 1 corresponds to the initial 2 CV of introduction of the Cleavage Buffer, while Cleavage Buffer 2 corresponds to the remaining 3 CV of introduction of the Cleavage Buffer; Elution 1 corresponds to the initial 2 CV of elution, while Elution 2 corresponds to the remaining 4 CV of elution; lastly, Stripping 1 corresponds to the initial 2.5 CV of stripping, while Stripping 2 corresponds to the remaining 3 CV of stripping. The grey line represents the step (0 to 100%) introduction of the Cleavage Buffer. The numbering below each sample phase is corresponding to the gel numbering. B) Polyacrylamide gel of the initial loaded sample and of the several collected fractions during chromatography. Lane 1: Initial loaded sample; lane 2: Sample loading flow through; lane 3: Column Wash 1; lane 4: Introduction of Cleavage Buffer 1; lane 5: Introduction of Cleavage Buffer 2; lane 6: Elution 1; lane 7: Elution 2; lane 8: Stripping 1; lane 9: Stripping 2. Before running the electrophoresis, some samples' concentration was firstly adjusted and the loaded amount in each well was also not constant, with a view to achieving a proper protein concentration in each lane.

Most of the conclusions for the present experiment are very similar to the previous ones, so only the main differences will be mentioned. Firstly, regarding the washing step using stringent washing conditions, Figure 19 B) (lane 3) shows that no significative improvement on the washing of nonspecific

bound proteins was achieved. In order to understand the effect of the up-flow elution, the first four 0.5 CV collected during elution were individually analyzed via an Enzyme α activity assay (data not shown). For the previous CTC01 down-flow elution experiment (section 4.3.1), 28%, 43%, 22% and 7% of Elution 1 activity was collected, respectively, in the first, second, third and fourth 0.5 CV of elution. For the present CTC01 up-flow elution experiment, 38%, 37%, 18% and 7% of Elution 1 was collected, respectively, in the first, second, third and fourth 0.5 CV of elution. These results show that up-flow elution indeed allowed to obtain a more concentrated elution fraction in the first 0.5 CV. The analysis of the chromatogram (Figure 19 A)) also shows, as expected, a clearer peak in the first collected fractions of elution phase.

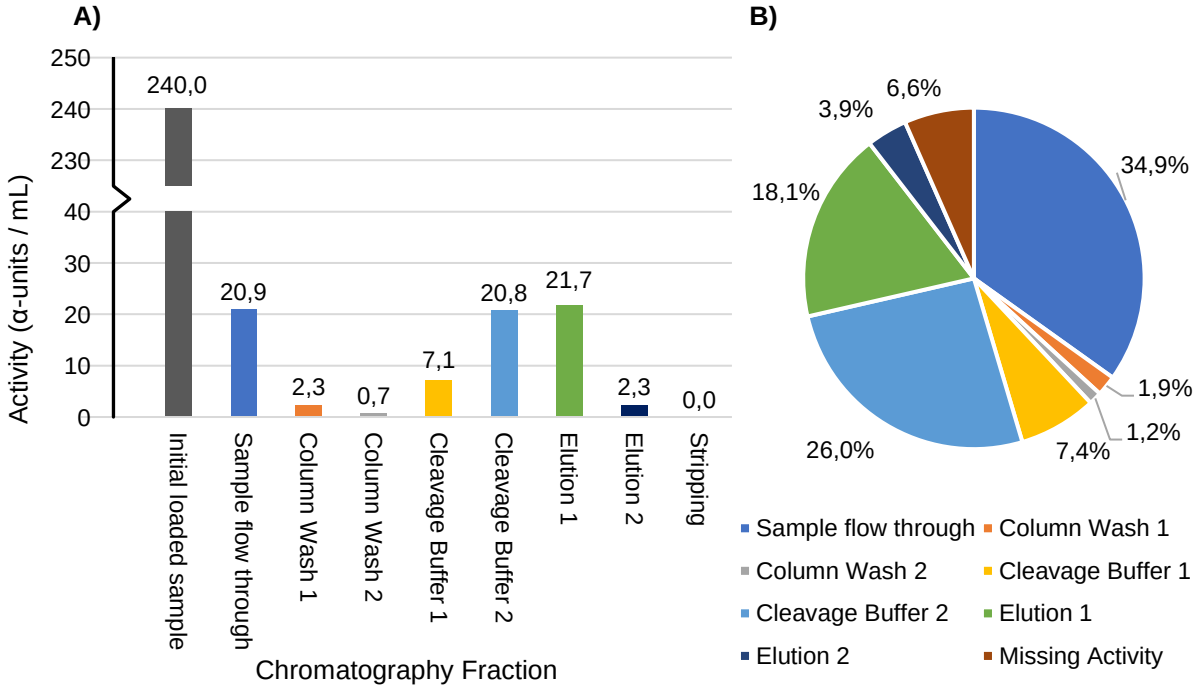

Figure 20 Enzyme α activity assay results after affinity purification of CTC01 following the "Affinity Chromatography – CBD v2" procedure. A) Activity measurements on the initial sample loaded into the column and on the whole fractions collected during chromatography. Note that it was not possible to measure activity on the Stripping phase due to the high pH that inactivates Enzyme α. B) Percentage of Enzyme α activity present in each fraction in relation to the total amount of Enzyme α activity initially loaded into the column. The total activity of Enzyme α in each fraction was calculated based on activity concentrations of A) and knowing the volume of each collected fraction. Assuming that there is no inactivation during the purification process and that both native Enzyme α and the fusion construct have the same molar catalytic activity (α-unit/mmol), the activity loaded into the column must be equal to the activity that leaves the column in the several different fractions, being possible to establish an Enzyme α activity balance to the affinity column. The missing activity was calculated in order to close the activity balance.

Figure 19 B) shows that the initial loaded sample (lane 1) is partially cleaved, being possible to estimate a 2:3 (fusion construct):(native Enzyme α) ratio. Taking this into account, and assuming an identical molar catalytic activity for CTC01 and for native Enzyme α, one could assume that at least 60% of the activity loaded into the column (corresponding to the native Enzyme α) should be washed from the column in the first phases. Figure 20 B) shows that 38.0% of the activity initially loaded into the column was washed out from the column during sample application and column wash, which is far below the expected value, possibly indicating that some native Enzyme α might show some kind of interaction with the stationary phase. However, when comparing the present experiment with the previous CTC01 down-

flow elution (section 4.3.1), the activity found in the fractions collected during the introduction of Cleavage Buffer were even higher (33.4%) suggesting, once again, that the introduction of the Cleavage Buffer, even before the 40 hours incubation, may already induce some premature cleavage of the fusion construct. Once again, note that during the introduction of the Cleavage Buffer there is an increase in the 280 nm absorbance (Figure 19 A)) most likely due to the presence of DTT in the Cleavage Buffer.

For the present case, assuming a 2:3 ratio (corresponding to $R = 0.4$) and using Equation 7, it was possible to calculate for Elution 1 fraction a $\eta_{Elution\ 1}^{purification} = 45\%$. Assuming that CTC01 chromatography could be optimized in order to avoid the premature washing of purified native Enzyme α during the introduction of the Cleavage Buffer, then it was possible to calculate for the combination of Introduction of Cleavage Buffer 2 and Elution 1 fractions a $\eta_{Cleavage\ Buffer\ 2\ +\ Elution\ 1}^{purification} = 110\%$. However, this yield higher than 100% shows that, in principle, the activity collected during Introduction of Cleavage Buffer 2 cannot fully correspond to fusion construct premature cleavage. Instead, the conjecture that this activity came from the combination of a premature construct cleavage with the washing of native Enzyme α already present in the initial sample seems the most plausible conclusion. Note that the purified cleaved Enzyme α present in both Introduction of Cleavage Buffer 2 and Elution 1 fractions were analyzed by mass spectrometry which confirmed that the cleavage occurred in the predicted cleavage site.

Lastly, an interesting difference is the much lower Enzyme α missing activity (6.6%) which, assuming no inactivation during chromatography, might indicate an almost complete cleavage of the fusion construct CTC01. Although lane 8 of Figure 19 B) indeed show the washing of the cleaved intein-tag (≈ 27 kDa) during stripping, no fusion construct was washed from the column. Instead, some native Enzyme α was still washed out as happened in the previous CTC01 down-flow experiment.

4.3.2. Fusion Construct 2 (CTC02)

Fusion Construct 2 (CTC02) combines both a chitin-binding domain (CBD) tag and a polyhistidine tag (his-tag) with the *Mxe* GyrA mini-intein to yield a self-cleaving purification tag. Enzyme α was genetically fused downstream of the tag, generating a fusion protein (CTC02) whose purification was tested through a chitin affinity chromatography and an immobilized metal affinity chromatography (IMAC). As already mentioned for CTC01, the *Mxe* GyrA intein N-terminal cleavage is induced by incubation with 50 mM DTT. The effect of the -1 residue (residue immediately preceding the splicing competent *Mxe* GyrA intein) on splicing and cleavage of the *Mxe* GyrA intein was studied in [35, 38] and, for the specific residues present in CTC02, the higher cleavage yield was also obtained during a 40 hours incubation at 23°C.

4.3.2.1. Chitin Affinity Chromatography

After packing a 1 mL bed volume of chitin resin into a Tricorn™ Column 5/50 and preparing the Intein Sample expressing CTC02 at 20°C, the purification of the fusion CTC02 was initially performed following the method “Affinity Chromatography – CBD v2” described in section 3.2.1.5. From the sample flow through during sample application, until stripping phase all the unbound material was collected in the

column flow through in 0.5 mL fraction volumes (0.5 CV) to be then analyzed via Enzyme α activity assay and NuPAGE® electrophoresis.

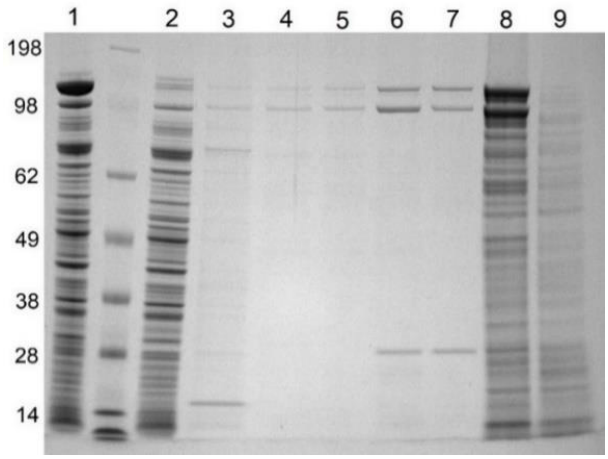

Figure 21 Polyacrylamide gel of the initial loaded sample and of the several collected fractions during chromatography. The fractions collected during column wash, introduction of the Cleavage Buffer, elution and stripping were split: Column Wash 1 corresponds to the initial 2 CV of column wash, while Column Wash 2 corresponds to the remaining 4 CV of column wash; Cleavage Buffer 1 corresponds to the initial 2 CV of introduction of the Cleavage Buffer, while Cleavage Buffer 2 corresponds to the remaining 3 CV of introduction of the Cleavage Buffer; Elution 1 corresponds to the initial 2 CV of elution, while Elution 2 corresponds to the remaining 4 CV of elution; lastly, Stripping 1 corresponds to the initial 3 CV of stripping, while Stripping 2 corresponds to the remaining 2.5 CV of stripping. Lane 1: Initial loaded sample; lane 2: Sample loading flow through; lane 3: Column Wash 1; lane 4: Introduction of Cleavage Buffer 1; lane 5: Introduction of Cleavage Buffer 2; lane 6: Elution 1; lane 7: Elution 2; lane 8: Stripping 1; lane 9: Stripping 2. Before running the electrophoresis, some samples' concentration was firstly adjusted and the loaded amount in each well was also not constant, with a view to achieving a proper protein concentration in each lane.

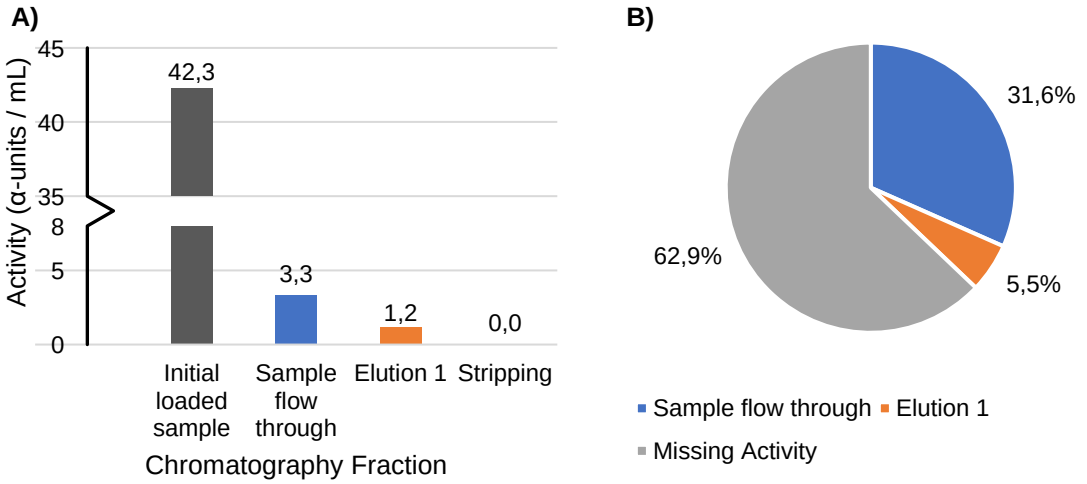

Figure 22 Enzyme α activity assay results after affinity purification of CTC02 following the "Affinity Chromatography – CBD v2" procedure. The fractions showing no activity were omitted from the graphics. A) Activity measurements on the initial sample loaded into the column and on some fractions collected during chromatography. Note that it was not possible to measure activity on the Stripping phase due to the high pH that inactivates Enzyme α. B) Percentage of Enzyme α activity present in each fraction in relation to the total amount of Enzyme α activity initially loaded into the column. The total activity of Enzyme α in each fraction was calculated based on activity concentrations of A) and knowing the volume of each collected fraction. Assuming that there is no inactivation during the purification process and that both native Enzyme α and the fusion construct have the same molar catalytic activity (α-unit/mmol), the activity loaded into the column must be equal to the activity that leaves the column in the several different fractions, being possible to establish an Enzyme α activity balance to the affinity column. The missing activity was calculated in order to close the activity balance.

As mentioned before, the comparison between the initial sample (Figure 21, lane 1) and the sample loading flow through (Figure 21, lane 2) shows that CTC02 bound to the chitin resin, while the native Enzyme α was simply washed out from the column.

Figure 21 shows that, as expected by the initial Fusion Constructs Screening (section 4.2), the initial loaded sample (lane 1) is partially cleaved, being possible to estimate a 3:1 (fusion construct):(native Enzyme α) ratio. Taking this into account, and assuming an identical molar catalytic activity for the CTC02 and for the native Enzyme α, one could assume that at least 25% of the activity loaded into the column (corresponding to the native Enzyme α) should be washed from the column in the first phases. Figure 22 B) shows that 31.6% of the activity initially loaded into the column was washed out from the column during sample application, which acceptably in accordance with the 3:1 ratio of the initial sample.

The purification yield for the purified sample was calculated based on Equation 7. For the present case, assuming a 3:1 ratio (corresponding to $R = 0.75$), it was possible to calculate for Elution 1 fraction a $\eta_{Elution\ 1}^{purification} = 7\%$. Associated to this very poor yield, Elution 1 fraction did also not show a promising purification factor (Figure 21, lane 6), and a significative amount of Enzyme α activity (62.9%) was missing in the activity balance. Although it was not possible to measure activity in the stripping phase due to the high pH (Figure 22 A)), Figure 21 (lanes 8 and 9) shows that there are proteins being washed from the column in this phase. In fact, as previously mentioned, after inducing the intein on-column cleavage and elute the cleaved Enzyme α, it is expected that some uncleaved fusion precursor and the cleaved intein-tag remain bound to the column. The stripping lanes of polyacrylamide gel indicate the washing of the fusion construct and native Enzyme α together with very small amounts of the cleaved intein-tag.

The aforementioned results lead to the conclusion that, for some unexpected reason, CTC02 cleavage did not occur. This conclusion is also supported by the fact that, as already discussed in section 4.2, this Intein Sample expressing fusion CTC02 showed a very low Enzyme α activity when compared to the other fusion constructs. Since CTC02 only differs from CTC01 for having an extra his-tag and for lacking a linker between Enzyme α and *Mxe* GyrA intein, one might conclude that either this linker, initially designed to promote a correct cleavage between Enzyme α and the intein, plays an important role in protein folding; or the his-tag affects the protein folding of the whole construct, consequently affecting its activity and cleavage properties. A possible justification could be the interaction of the his-tag with the metals present in the stabilizing buffer, or the subtraction of these metals from Enzyme α.

4.3.2.2. Immobilized Metal Affinity Chromatography (IMAC)

Even though the previous chitin affinity chromatography with CTC02 did not show very promising results, an IMAC purification was still performed using a 1 mL HiTrap™ TALON® crude prepacked with TALON Superflow™, after preparation of the Intein Sample expressing CTC02 at 20°C.

The analysis of Figure 23 revealed that all the initial sample loaded into the column (lane 1) left the column during sample loading flow through (lane 2), indicating that no binding of CTC02 to the affinity resin occurred. This conclusion is also supported by the activity assays results (data not shown) which

confirmed that 100% of the activity loaded into the column was collected in the sample loading flow through fractions.

The unsuccessful IMAC purification of CTC02 suggests that either the his-tag was not expressed on the fusion construct, or it may be completely inaccessible due to an unexpected folding. Further conclusions regarding this CTC02 were already discussed in section 4.3.2.1.

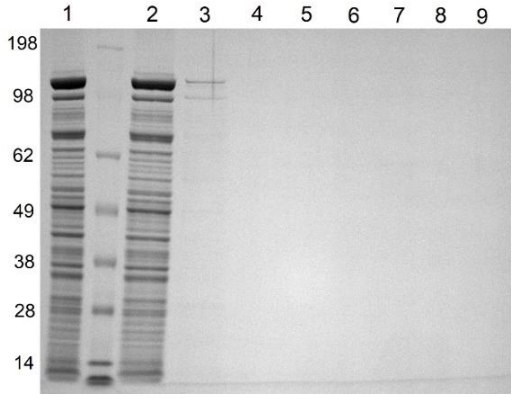

Figure 23 Polyacrylamide gel of the initial loaded sample and of the several collected fractions during chromatography. The fractions collected during column wash, introduction of the Cleavage Buffer, elution and stripping were split: Column Wash 1 corresponds to the initial 4 CV of column wash, while Column Wash 2 corresponds to the remaining 11 CV of column wash; Cleavage Buffer 1 corresponds to the initial 3 CV of introduction of the Cleavage Buffer, while Cleavage Buffer 2 corresponds to the remaining 2 CV of introduction of the Cleavage Buffer; Elution 1 corresponds to the initial 3 CV of elution, while Elution 2 corresponds to the remaining 3 CV of elution; lastly, Stripping 1 corresponds to the initial 3 CV of stripping, while Stripping 2 corresponds to the remaining 5 CV of stripping. Lane 1: Initial loaded sample; lane 2: Sample loading flow through; lane 3: Column Wash 1; lane 4: Introduction of Cleavage Buffer 1; lane 5: Introduction of Cleavage Buffer 2; lane 6: Elution 1; lane 7: Elution 2; lane 8: Stripping 1; lane 9: Stripping 2.

4.3.3. Fusion Construct 3 (CTC03)

Fusion Construct 3 (CTC03) combines the chitin-binding domain (CBD) tag with the *Mtu* RecA mini-intein to yield a self-cleaving purification tag. Enzyme α was genetically fused downstream of the tag, generating a fusion protein (CTC03) whose purification was tested using a chitin resin. In theory, the intein is stable at pH=8.5, while its C-terminal cleavage is induced by a 20 to 24 hours incubation at pH=6 to 6.5 and 25°C [29, 34, 37, 59, 60, 61]. The intein self-cleavage reaction is also known to be temperature dependent and, consequently, the reaction can also be carried out at 37°C for 4 to 6 hours [37] or for longer periods of 14 to 18 hours [62]. Although a shorter incubation time is preferred, the column incubation at 37°C was not viable for the present feasibility study and, for that reason, an incubation at room temperature was performed for 22.5 hours. Also, in order to avoid interfering with Enzyme α stability, the intein cleavage was induced by a mild pH shifting from 8 to 6.5.

After packing a 1 mL bed volume of chitin resin into a Tricorn™ Column 5/50 and preparing the Intein Sample expressing CTC03 at 20°C, the purification of the fusion CTC03 was initially performed following the method “Affinity Chromatography – CBD v1” described in section 3.2.1.5. From the sample flow through during sample application, until stripping phase all the unbound material was collected in the column flow through in 0.5 mL fraction volumes (0.5 CV) to be then analyzed via Enzyme α activity assay and NuPAGE® electrophoresis.

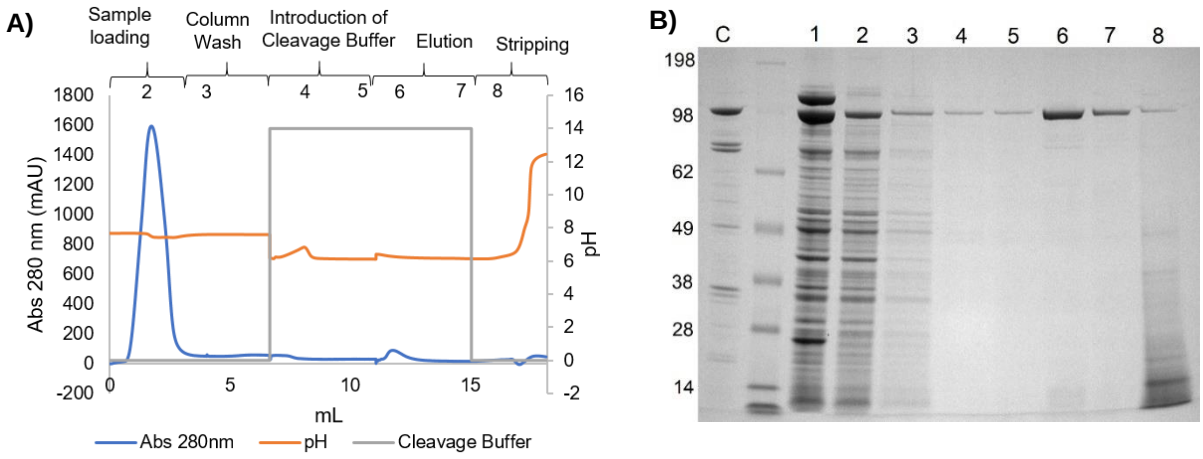

Figure 24 A) Chromatogram of CTC03 purification following "Affinity Chromatography – CBD v1". The fractions collected during column wash, introduction of the Cleavage Buffer, elution and stripping were split: Column Wash 1 corresponds to the initial 2 CV of column wash, while Column Wash 2 corresponds to the remaining 4 CV of column wash; Cleavage Buffer 1 corresponds to the initial 2 CV of introduction of the Cleavage Buffer, while Cleavage Buffer 2 corresponds to the remaining 3 CV of introduction of the Cleavage Buffer; Elution 1 corresponds to the initial 2 CV of elution, while Elution 2 corresponds to the remaining 4 CV of elution; lastly, Stripping 1 corresponds to the initial 2.5 CV of stripping, while Stripping 2 corresponds to the remaining 3 CV of stripping. The grey line represents the step (0 to 100%) introduction of the Cleavage Buffer accompanied by a decrease in the pH. The numbering below each sample phase is corresponding to the gel numbering. B) Polyacrylamide gel of the purified Enzyme α control sample, the initial loaded sample and of the several collected fractions during chromatography. Lane C: Enzyme α control purified sample by ion exchange chromatography; lane 1: Initial loaded sample; lane 2: Sample loading flow through; lane 3: Column Wash 1; lane 4: Introduction of Cleavage Buffer 1; lane 5: Introduction of Cleavage Buffer 2; lane 6: Elution 1; lane 7: Elution 2; lane 8: Stripping 1. Before running the electrophoresis, some samples' concentration was firstly adjusted and the loaded amount in each well was also not constant, with a view to achieving a proper protein concentration in each lane.

Firstly, the comparison between the initial sample (Figure 24 B), lane 1) and the sample loading flow through (Figure 24 B), lane 2) shows that CTC03 did bind to the chitin resin, while the native Enzyme α, for not having a CBD, was simply washed out from the column.

Figure 24 B) shows that the initial loaded sample (lane 1) is partially cleaved with a 1:1 (fusion construct):(native Enzyme α) ratio. Assuming an identical molar catalytic activity for the CTC03 and for the native Enzyme α, one might assume that at least 50% of the activity loaded into the column (corresponding to the native Enzyme α) should be washed in the first phases. Figure 25 B) shows that 47.2% of the activity initially loaded into the column was washed out during sample application and column washing, which agrees with the 1:1 ratio of the initial sample.

Regarding purity, the fractions obtained during Elution 1 represented in lane 6 of the polyacrylamide gel did qualitatively show a very high purification factor. An Enzyme α purified sample by ion exchange (IEX) chromatography was included in lane C for comparison. The clearly higher purity achieved with chitin affinity chromatography enhances the potential of this technique. The purified cleaved Enzyme α present in Elution 1 fractions was analyzed by mass spectrometry which confirmed that the cleavage occurred in the predicted cleavage site.

The purification yield for the purified sample was calculated based on Equation 7. For the present case, assuming a 1:1 ratio (corresponding to $R = 0.5$), it was possible to calculate for Elution 1 fraction a $\eta_{Elution\ 1}^{purification} = 96\%$.

To close the activity balance, only an insignificant Enzyme α activity was missing (0.5%), which indicates a complete cleavage of the fusion constructs. However, upon stripping no band corresponding to the intein-tag was observed in lane 8 of Figure 24 B).

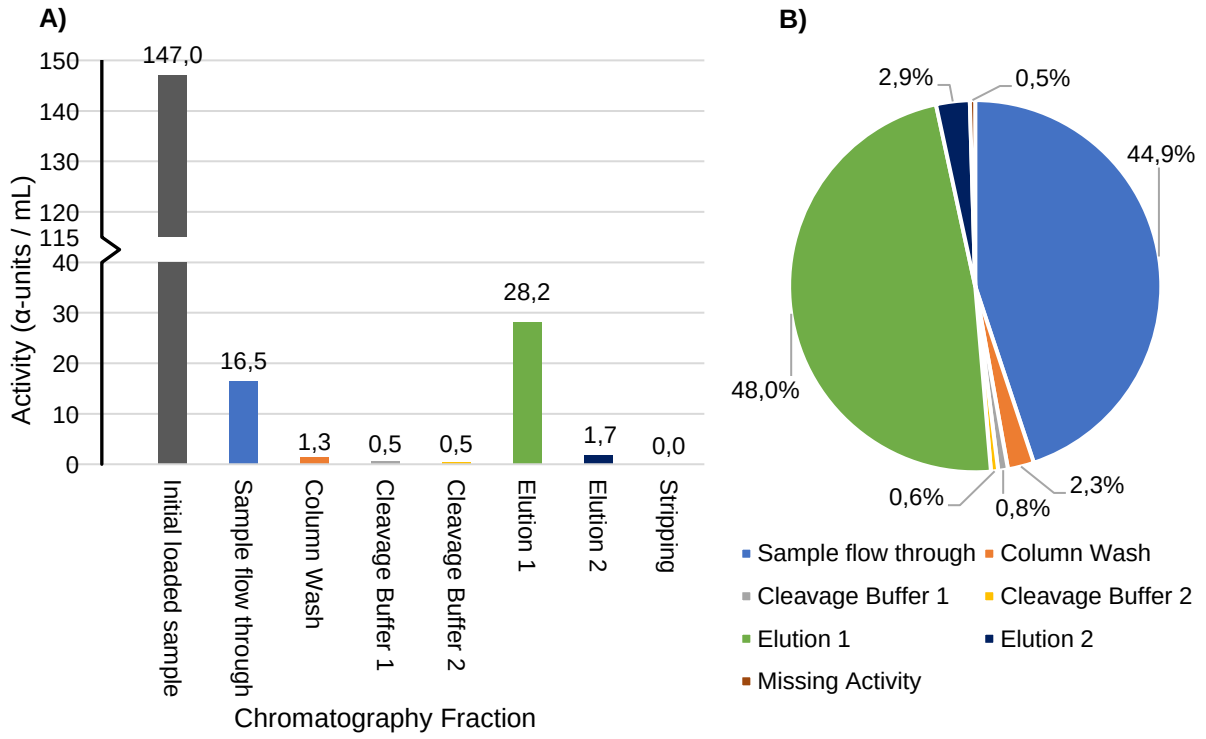

Figure 25 Enzyme α activity assay results after affinity purification of CTC03 following the "Affinity Chromatography – CBD v1" procedure. A) Activity measurements on the initial sample loaded into the column and on the whole fractions collected during chromatography. Note that it was not possible to measure activity on the Stripping phase due to the high pH that inactivates Enzyme α. B) Percentage of Enzyme α activity present in each fraction in relation to the total amount of Enzyme α activity initially loaded into the column. The total activity of Enzyme α in each fraction was calculated based on activity concentrations of A) and knowing the volume of each collected fraction. Assuming that there is no inactivation during the purification process and that both native Enzyme α and the fusion construct have the same molar catalytic activity (α-unit/mmol), the activity loaded into the column must be equal to the activity that leaves the column in the several different fractions, being possible to establish an Enzyme α activity balance to the affinity column. The missing activity was calculated in order to close the activity balance.

4.4. Affinity Chromatography with Fusion Construct 3 (CTC03) Sample

After analyzing the previous results with the Intein Samples expressing CTC01, CTC02 and CTC03, it is possible to conclude that both CTC01 and CTC03 are very promising constructs to study in depth. In fact, they allowed the purification of Enzyme α with very optimistic yields and purities. However, the affinity chromatography of CTC01 still needs some optimization in order to avoid the suspected early cleavage of the fusion construct during the introduction of the Cleavage Buffer. On the other hand, the purification of CTC03 showed a very positive result and, for that reason, a new purification attempt with a higher sample loading volume into the column was performed with Fusion Construct 3 Sample. This sample is also expressing CTC03 and only differs from the previous tested Intein Sample expressing CTC03 in its production protocol (section 3.2.1.3). In order to have an efficient higher volume sample

loading, this sample was concentrated four times in comparison with the previous Intein Sample expressing CTC03.

4.4.1. Dynamic Binding Capacity (DBC) Test

As the main goal of the present project lays on the concept of enzyme chromatography purification by adsorbing the fusion construct and wash all the impurities in the flow through, it is important to determine the volume of feedstock possible to load into the column before it gets saturated in fusion construct and it starts to appear in the flow through. To exactly determine a representative dynamic binding capacity of the column, a pure fusion sample should be used. However, since such a sample was not available, this DBC test was performed with the Fusion Construct 3 Sample. This Fusion Construct 3 Sample also showed an early cleavage with a 1:1 (fusion construct):(native Enzyme α) ratio and the presence of the cleaved intein-tag (≈ 27 kDa). This cleaved intein-tag is expected to interact with the column and hence decrease its binding capacity. Therefore, the DBC for the intact fusion should be dependent on the extent of premature cleavage.

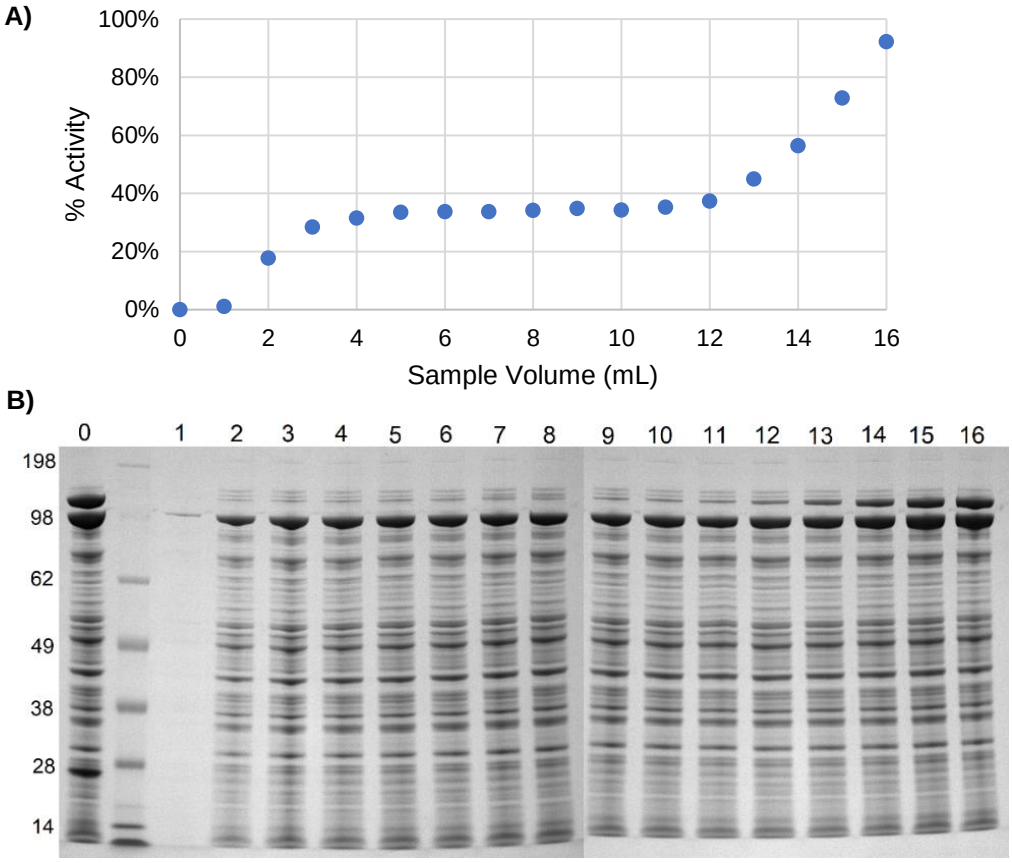

Figure 26 A) Breakthrough curve for Fusion Construct 3 Sample as function of loaded sample volume. Y-axis refers to the percentage ratio between the measured activity in the flow through and the measured activity in the initial sample. 16 mL of sample were loaded into a Tricorn 5/50 packed with a 1 mL bed volume of chitin resin. B) Polyacrylamide gel of the collected unbound material in the flow through. Lane 0 represents the initial loaded sample, lane 1 the flow through after 1 mL loading, lane 2 the flow through after loading the 2nd mL, lane 3 the flow through after loading the 3rd mL, until lane 16 after loading the 16th mL.

As already discussed in section 4.3.1.1, based on the theoretical binding capacity provided by the chitin resin manufacturer [38], it was possible to estimate a binding capacity of around 11 mg of Enzyme α per

mL of bed volume. After production, lysis and preparation of the Fusion Construct 3 (CTC03) sample, an Enzyme α activity assay was performed on this initial sample to determine its Enzyme α concentration. Considering an identical molar catalytic activity (α-unit/mmol) between the fusion and the native Enzyme α, and assuming the cleavage ratio of 1:1, it was estimated that at least 8.5 mL of sample should be loaded into the column to achieve the theoretical maximum binding capacity. In order to be completely sure that the full column capacity would be reached, 16 mL of this initial sample was applied into the column, being the unbound material in the flow through collected and analyzed via NuPAGE® electrophoresis and Enzyme α activity assay.

The breakthrough curve represented in Figure 26 A) showed an expected trend, with a first step corresponding to the measured activity of the native Enzyme α (without a CBD due to the early cleavage) which is immediately washed from the column. It also shows a second step in which activity started to increase due to the elution of the fusion construct at the moment of resin saturation. After loading between 11 to 12 mL of Fusion Construct 3 Sample, it is possible to detect both an increase in activity (Figure 26 A)) and an increase in the intensity of the band associated to the fusion construct (Figure 26 B)), suggesting the moment of column saturation. After 16 mL sample loading the eluted fractions showed almost 100% of the activity of the initial sample, which indicates a full column saturation.

Two remarks must be added regarding the DBC results. Firstly, assuming a 1:1 (fusion construct):(native Enzyme α) ratio in the initial sample and an identical molar catalytic activity (α-unit/mmol) for both the native Enzyme α and the fusion construct, one was expecting that the first step in Figure 26 A) would be closer to the 50% of the measured activity in the initial sample. The fact that it is below the expected value, might indicate that some native Enzyme α might also be retained in the column. Secondly, it was also expected that upon column saturation, both the fusion construct and the cleaved intein-tag present in the initial sample would start to be washed from the column. This indeed happened to the fusion construct, but the washing of the cleaved intein-tag is not evident. This may indicate that this tag either has a higher affinity for the resin than CTC03 and might compete for the binding, or that for its smaller size it is able to bind to the resin in some positions where a fusion construct cannot fit.

4.4.2. Chitin Affinity Chromatography

After the dynamic binding capacity test, the purification of the Fusion Construct 3 (CTC03) was initially performed following the method “Affinity Chromatography – CBD v3” described in section 3.2.1.5. “Affinity Chromatography CBD - v3” was the last performed purification, being similar to “Affinity Chromatography CBD v1”, but with a few differences mainly in the sample loading step. In this last protocol, considering the results of the DBC study, the column was loaded with 7.5 mL of Fusion Construct 3 Sample (using around 70% of the column capacity), in order to understand the influence of higher volume sample loading in the course of the chromatography, and to try to obtain a more concentrated purified Enzyme α sample. From the sample flow through during sample application, until stripping phase all the unbound material was collected in the column flow through in 0.5 mL fraction volumes (corresponding to 0.5 CV) to be then analyzed via Enzyme α activity assay and NuPAGE® electrophoresis.

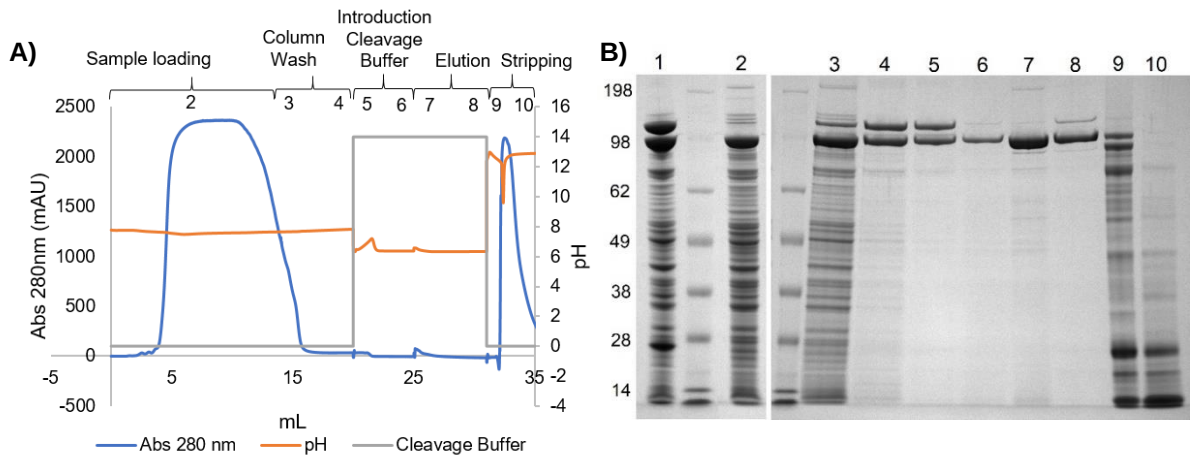

Figure 27 A) Chromatogram of CTC03 purification following "Affinity Chromatography – CBD v3". The fractions collected during column wash, introduction of the Cleavage Buffer, elution and stripping were split: Column Wash 1 corresponds to the initial 3 CV of column wash, while Column Wash 2 corresponds to the remaining 3 CV of column wash; Cleavage Buffer 1 corresponds to the initial 2 CV of introduction of the Cleavage Buffer, while Cleavage Buffer 2 corresponds to the remaining 3 CV of introduction of the Cleavage Buffer; Elution 1 corresponds to the initial 2 CV of elution, while Elution 2 corresponds to the remaining 4 CV of elution; lastly, Stripping 1 corresponds to the initial 2.5 CV of stripping, while Stripping 2 corresponds to the remaining 3 CV of stripping. The grey line represents the step (0 to 100%) introduction of the Cleavage Buffer accompanied by a decrease in the pH. The numbering below each sample phase is corresponding to the gel numbering. B) Polyacrylamide gel of the initial loaded sample and of the several collected fractions during chromatography. Lane 1: Initial loaded sample; lane 2: Sample loading flow through; lane 3: Column Wash 1; lane 4: Column Wash 2; lane 5: Introduction of Cleavage Buffer 1; lane 6: Introduction of Cleavage Buffer 2; lane 7: Elution 1; lane 8: Elution 2; lane 9: Stripping 1; lane 10: Stripping 2. Before running the electrophoresis, some samples' concentration was firstly adjusted and the loaded amount in each well was also not constant, with a view to achieving a proper protein concentration in each lane.

Firstly, comparing the initial sample (Figure 27 B), lane 1) with the sample loading flow through (Figure 27 B), lane 2) it is possible to observe the binding of the CTC03, while the native Enzyme α, lacking a CBD, was washed from the column. These two bands were analyzed by mass spectrometry which confirmed their identity. The more intense band with a MW ≈ 25 kDa in the initial sample was also analyzed by MS which confirmed to be the cleaved intein-tag.

Figure 27 B) shows that the initial loaded sample (lane 1) is partially cleaved with an estimated 1:1 (fusion construct):(native Enzyme α) ratio. Taking this into account, and assuming an identical molar catalytic activity for CTC03 and for native Enzyme α, one could assume that at least 50% of the activity loaded into the column (corresponding to the native Enzyme α) should be washed from the column in the first phases. Figure 28 B) shows that only 37% of the activity initially loaded into the column is washed out during sample application and column wash, which is below the expected value, indicating that not all native Enzyme α was washed from the column. Note that this result is in accordance with the formulated hypothesis in the previous dynamic binding capacity test (section 4.4.1) which also suggested that some native Enzyme α might be retained in the column together with the fusion CTC03.

Regarding purity, the fractions obtained during Elution 1 represented in lane 7 of the polyacrylamide gel, did qualitatively show a very high purification factor. However, the purification yield for the purified sample was calculated based on Equation 7 and, assuming a 1:1 ratio (corresponding to $R = 0.5$), a very poor $\eta_{Elution\ 1}^{purification} = 2\%$ was obtained for Elution 1 fraction. Even considering the combination of Elution 1 with Elution 2 fraction (which also showed a high purity), the calculated purification yield is still

very poor ($\eta^{purification}_{Elution\ 1+Elution\ 2} = 3.2\%$). The purified cleaved Enzyme α present Elution 1 fraction was analyzed by mass spectrometry which confirmed that the cleavage occurred in the predicted cleavage site.

The majority of Enzyme α activity (60.3%) belongs to the missing activity in the activity balance, which indicates that the most reasonable explanation for such low purifications yields is the poor cleavage of the fusion construct. After the previous experiments, this was a completely unexpected result and the reason behind the previous successful purification of the CTC03 Intein Sample (section 4.3.3), and the unsuccessful purification of the present Fusion Construct 3 Sample is not completely understood.

An interesting final remark is that the missing activity converted to Enzyme α mass (using its specific activity) corresponds to 10.5 mg of Enzyme α. Assuming no inactivation during the chromatography, this 10.5 mg corresponds to the amount of Enzyme α that was kept inside the column after elution phase. Interestingly, this value is close to the theoretical MBP binding capacity (provided by the chitin resin manufacturer [38]) extrapolated to Enzyme α (already discussed in section 4.3.1.1). This last evidence supports the previous hypothesis that no cleavage occurred in this purification experiment.

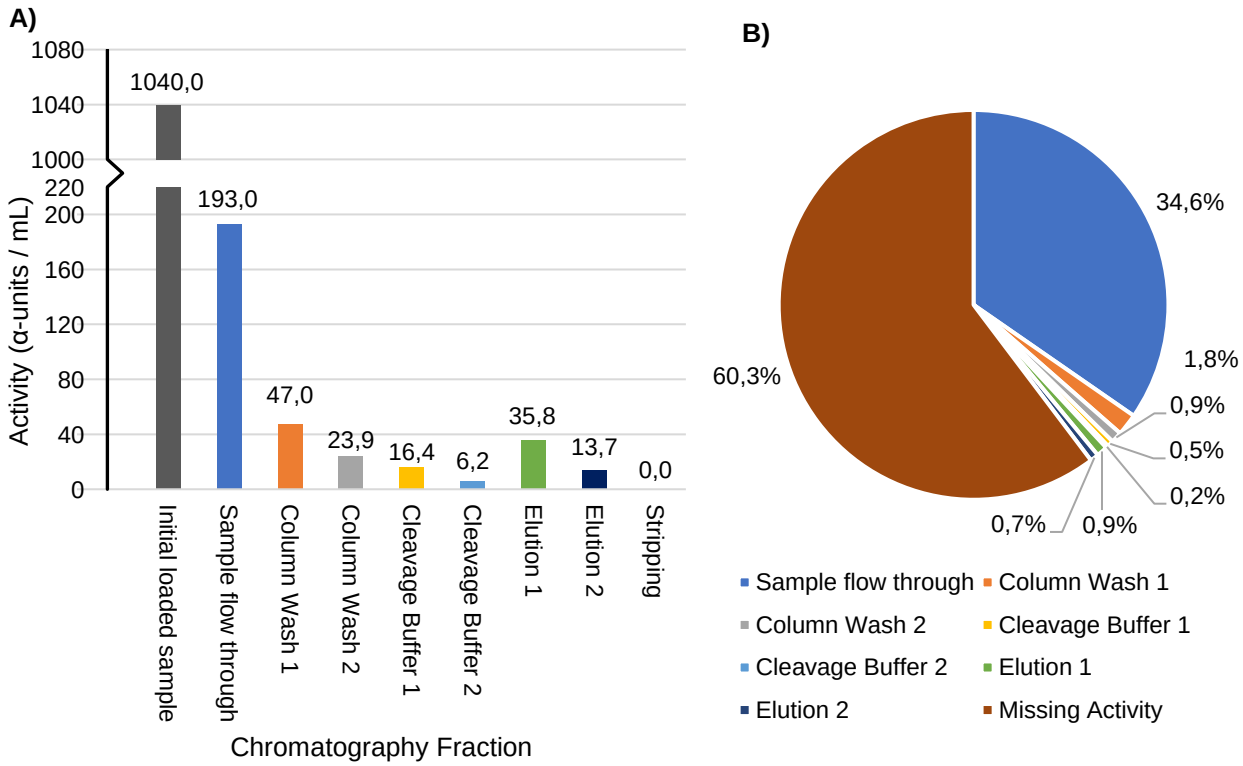

Figure 28 Enzyme α activity assay results after affinity purification of CTC03 following the "Affinity Chromatography – CBD v3" procedure. A) Activity measurements on the initial sample loaded into the column and on the whole fractions collected during chromatography. Note that it was not possible to measure activity on the Stripping phase due to the high pH that inactivates Enzyme α. B) Percentage of Enzyme α activity present in each fraction in relation to the total amount of Enzyme α activity initially loaded into the column. The total activity of Enzyme α in each fraction was calculated based on activity concentrations of A) and knowing the volume of each collected fraction. Assuming that there is no inactivation during the purification process and that both native Enzyme α and the fusion construct have the same molar catalytic activity (α-unit/mmol), the activity loaded into the column must be equal to the activity that leaves the column in the several different fractions, being possible to establish an Enzyme α activity balance to the affinity column. The missing activity was calculated in order to close the activity balance.

4.5. Batch Cleavage Test

After performing the chitin affinity chromatography with the Fusion Construct 3 Sample and realizing that fusion constructs cleavage might not be as reproducible as anticipated, a batch cleavage test was performed for the Intein Samples expressing CTC01, CTC02, CTC03 and CTC04 as well as for the Fusion Construct 3 Sample (in principle, identical to the Intein Sample expressing CTC03). These samples were incubated off-column, in the same Cleavage Buffers and under identical cleavage conditions of the previous on-column purifications: 40 hours in 50 mM DTT for CTC01 and CTC02 and 22.5 hours at pH=6.50 for CTC03 and CTC04.

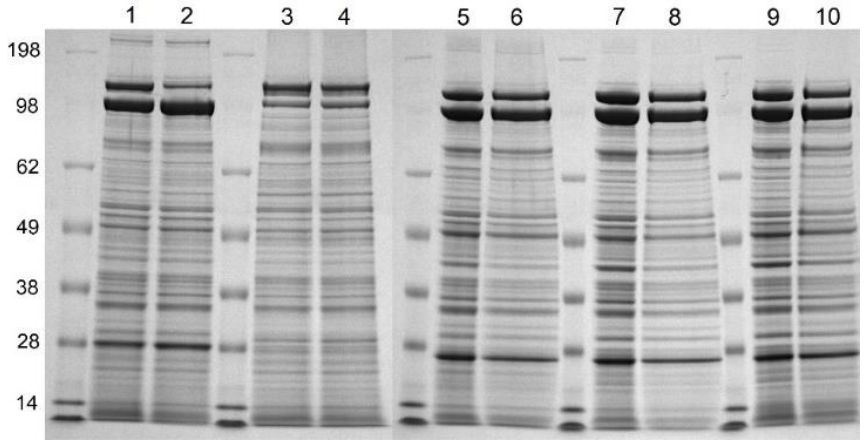

Figure 29 Polyacrylamide gel with both the initial samples and the samples after off-column cleavage induction. Lane 1: CTC01 initial Intein Sample; lane 2: CTC01 Intein Sample after cleavage induction; lane 3: CTC02 initial Intein Sample; lane 4: CTC02 Intein Sample after cleavage induction; lane 5: CTC03 initial Intein Sample; lane 6: CTC03 Intein Sample after cleavage induction; lane 7: initial Fusion Construct 3 Sample; lane 8: Fusion Construct 3 Sample after cleavage induction; lane 9: CTC04 initial Intein Sample; lane 10: CTC04 Intein Sample after cleavage induction. Before running the electrophoresis, some samples' concentration was firstly adjusted and the loaded amount in each well was also not constant, with a view to achieving a proper protein concentration in each lane.

The comparison between the initial samples' lanes with the cleaved samples' lanes of Figure 29 shows that only CTC01 did undergo a clear cleavage reaction. The fact that both the Intein Sample expressing CTC02 and the Fusion Construct 3 (CTC03) Sample did not show a proper cleavage, is in accordance with the obtained purification results for both constructs (sections 4.3.2 and 4.4.2). However, the observation that Intein Sample expressing CTC03 did not cleave in this batch test (Figure 29, lanes 5 and 6) is in contradiction with the purification results achieved with this sample in section 4.3.3. This contradiction could be explained by a different response from the fusion constructs to cleavage when they are bound to a chitin resin (on-column cleavage) and when they are in solution (off-column cleavage).

4.6. Molecular Modelling

4.6.1. Native Enzyme α

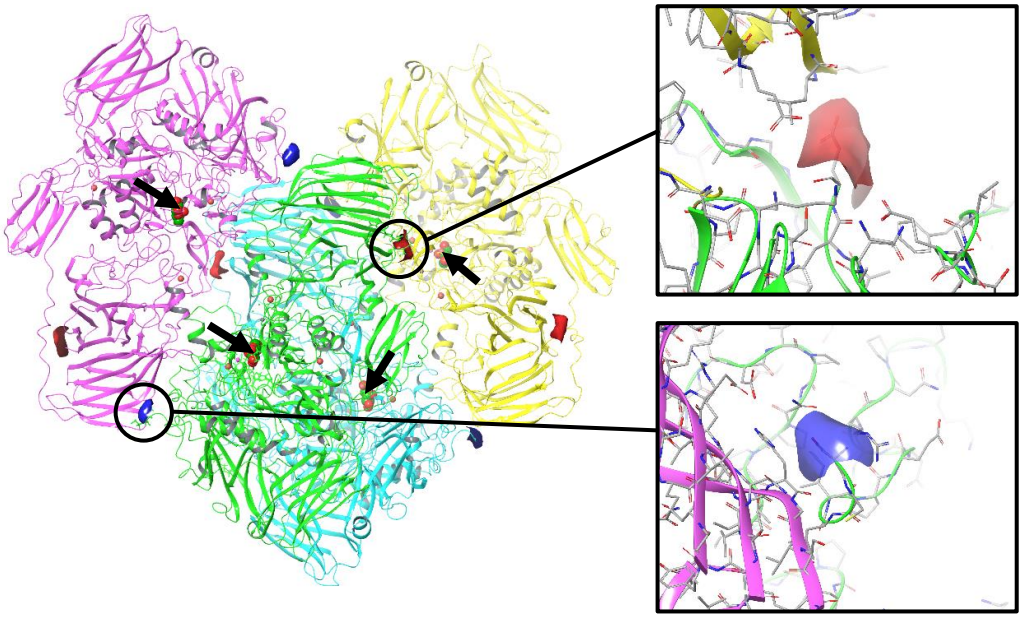

Figure 30 Enzyme α structure formed by four monomers (represented in green, purple, blue and yellow). The available surface of N and C-terminus of each monomer is represented, respectively, in blue and red. Both termini of the monomer represented in green are zoomed. The active site of each monomer is indicated with a black arrow.

Figure 30 shows that, in principle, the attachment of an intein-tag to either N or C-terminus of each monomer is possible due to the significant surface availability, even in the context of the tetramer. In fact, looking to the zoomed N and C-terminus of the monomer represented in green, it possible to identify a free region for the intein-tag attachment without disturbing the overall structure of Enzyme α.

4.6.2. Fusion Constructs

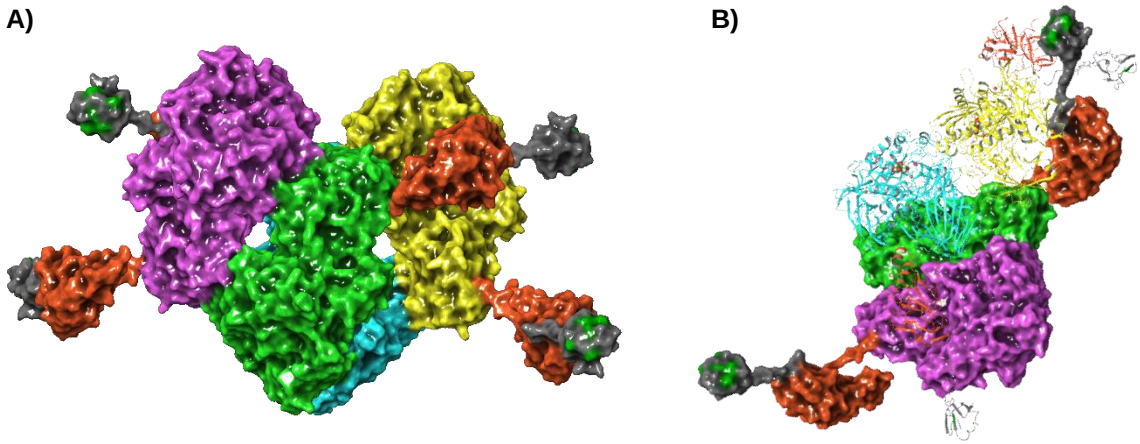

Figure 31 Fusion Construct 1 (CTC1) tetramer structure. The four monomers of Enzyme α are represented in green, purple, blue and yellow. The *Mxe* GyrA mini-intein attached to each monomer is represented in brown. Each CBD is represented in grey and contains some green areas, representing the residues known to interact with the chitin resin. A) Surface model for CTC01 tetramer. B) Model of CTC01 tetramer, with one of the dimers represented in ribbons and the other with its surface.

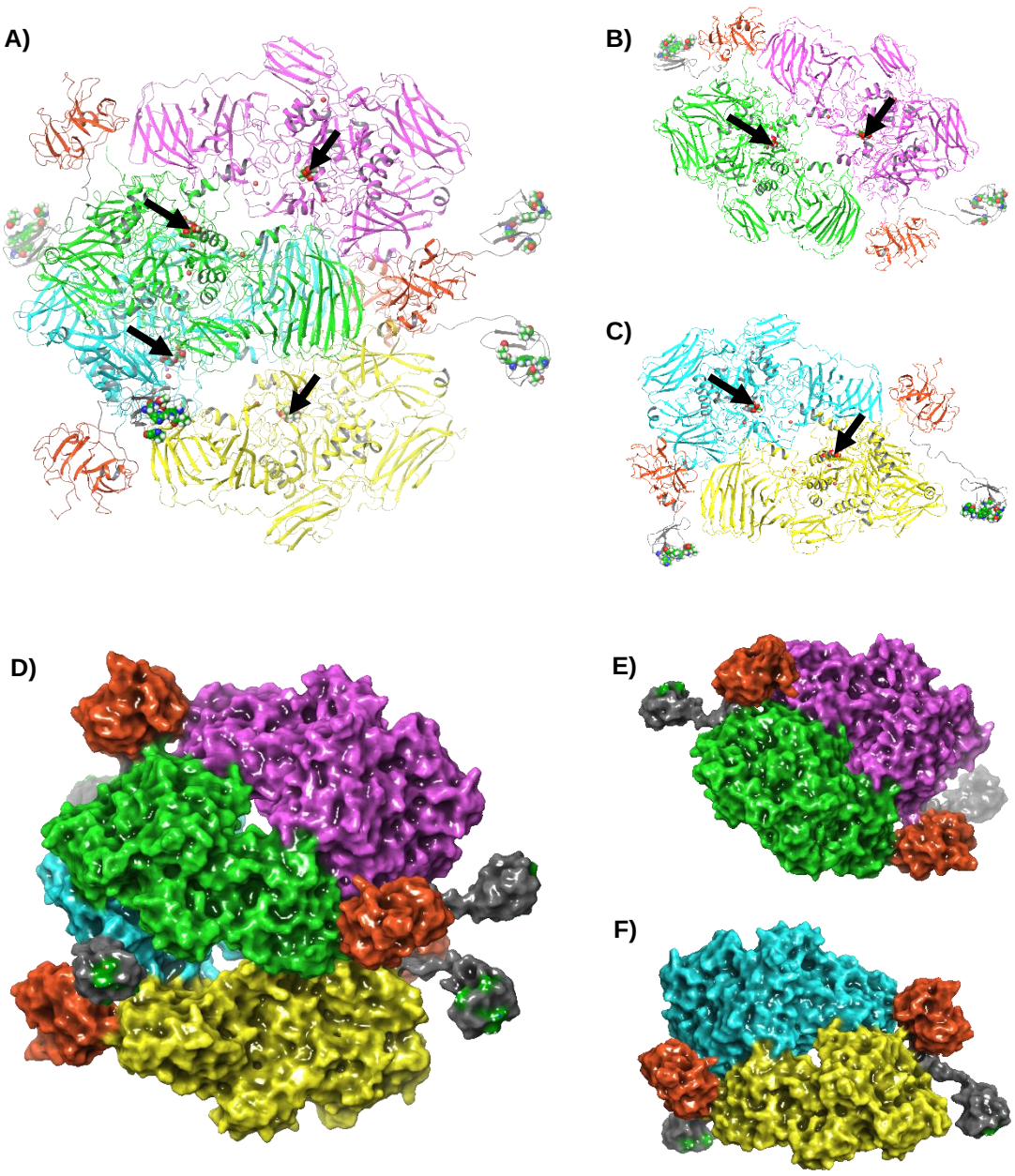

Figure 32 Fusion Construct 3 (CTC03) tetramer structure. The four monomers of Enzyme α are represented in green, purple, blue and yellow. The *Mtu* RecA mini-intein attached to each monomer is represented in brown. Each CBD is represented in grey and contains some green areas, representing the residues known to interact with the chitin resin. A) Ribbons model for the CTC03 tetramer. Each tetramer is formed by the association of two dimers represented in B) and C), being each active site represented with a black arrow. D) Surface model for the CTC03 tetramer. Each tetramer is formed by the association of two dimers represented in E) and F).

The analysis of Figures 31 and 32 suggests that, in principle, the expression of the fusions CTC01 and CTC03 should not result in any problem. For both constructs there is enough space for the intein-tag attachment to either the N or C-terminus of Enzyme α. There is some uncertainty in the positioning of the intein and CBD due to the flexible linkers. Still, it is unlikely that the intein-tag extensions will hamper access to the active site and, for that reason, the molar catalytic activity (α-unit/mmol) between the native Enzyme α and each fusion construct should not differ. Also, worth to mention is the fact that the chitin binding amino acids are all exposed and available for binding to the chitin resin, which is in accordance with the positive binding results in the previous purification experiments. Lastly, the analysis

of Figure 33 allows to conclude that, in principle, both mini-inteins present in CTC01 and CTC03 do have a spatial conformation that allows the respective intein N-terminal and C-terminal cleavage.

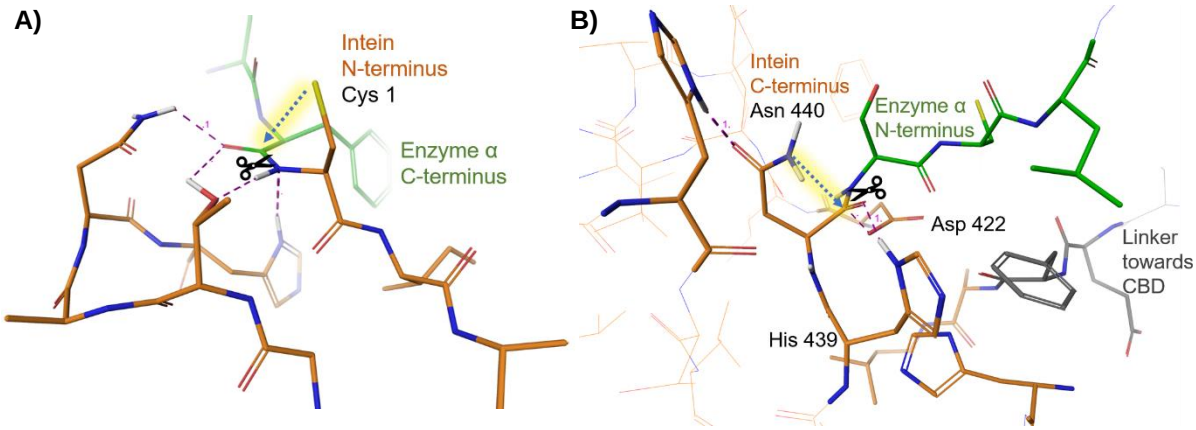

Figure 33 CTC01 and CTC03 cleavage sites details. The green, brown and grey chains represent, respectively, the Enzyme α, the mini-intein and the CBD. A) CTC01 details of the cleavage site at the *Mxe* GyrA mini-intein N-terminus. The intein was modified by an asparagine (Asn) to alanine (Ala) mutation in its C-terminus which inactivates both splicing and C-terminal cleavage activities (not visible). Intein N-terminal cleavage occurs by a N-S acyl-shift (indicated by a yellow arrow), forming a thioester linkage between Enzyme α and the intein (the broken bond is indicated with a black scissors). B) CTC03 details of the cleavage site at the *Mtu* RecA mini-intein C-terminus. The intein was modified by a cysteine (Cys) to alanine (Ala) mutation in its N-terminus (C1A) to enhance the suppression of N-terminal cleavage. Intein C-terminal cleavage is mediated by the cyclization of a C-terminal asparagine residue (Asn440) indicated by a yellow arrow (the broken bond is indicated with a black scissors). Some of the conserved residues are also shown.

5. CONCLUSIONS AND FUTURE PROSPECTS

The aim of the present project was to evaluate the feasibility of an affinity chromatography purification taking advantage of inteins as self-cleaving tags.

The initial screening of the six different fusion constructs indicated that, under the same fermentation conditions, it may not be possible to achieve comparable expression levels between fusions expressing different POI. In fact, even though for fusion constructs expressing Enzyme α the expression levels were very positive, the same did not happen for the fusion constructs expressing Enzyme β. This suggests that a "one fits all" solution is not possible and that different fusion constructs should be tested with additional linkers.

After expression and cell lysis, all the fusion samples having Enzyme α as the POI revealed the presence of the corresponding fusion construct (as expected), but also the presence of native Enzyme α (not expected). The presence of native Enzyme α may have two different origins: either, for some unexpected reason, native Enzyme α was individually expressed during sample production, or the expressed fusion construct underwent early cleavage. Considering the presence of the cleaved intein-tag in the initial samples, together with the already discussed cleavage conditions for each intein, the most reasonable hypothesis should be the premature cleavage during fermentation or cell lysis.

Having a partially cleaved initial sample was indeed a problem that raised difficulties on the whole subsequent purification process. Firstly, activity assays do not distinguish the native Enzyme α from the fusion constructs, complicating the Enzyme α tracking. Secondly, although there was no clear evidence that the fusion constructs showed a different molar catalytic activity from the native Enzyme α, it is still a possibility that should be kept in mind. Note that assuming an identical molar catalytic activity (α-unit/mmol) between these two structures, due to the higher molecular weight of the fusion construct, this should show a lower specific activity (α-unit/mg) when compared to the native Enzyme α. Lastly, after inducing the cleavage of the fusion constructs, it was completely impossible to distinguish the cleaved Enzyme α by the Cleavage Buffer, from the remaining native Enzyme α which was already present in the initial sample and which was not removed from the purification system.

For these reasons, future work should focus on the initial sample optimization in order to only express the intact fusion constructs. That would not only facilitate the interpretation of the following purifications results but would also be a critical step in achieving decent purification yields. In principle, this optimization should be performed both at the fermentation and cell lysis phase, in order to understand where the fusion is being prematurely cleaved. Also, it would be interesting to optimize the expression of fusions with different POI to understand whether this premature cleavage is induced by some fermentation or cell lysis conditions, or if the POI itself may have some influence on it. Regarding fermentation, the present project showed that the temperature at which protein expression is performed has a significative influence in both expression levels and (fusion construct):(native POI) ratios. For this reason, different expression temperatures (especially lower temperatures) and incubation times should be tested in order to achieve optimal expression levels and (fusion construct):(native POI) ratios. In

addition, the testing of different inteins and its further optimization by protein engineering, as well as the testing of different expression hosts may also provide relevant results.

Regarding the purification experiments with the Intein Samples expressing CTC01 and CTC03, both had very promising results, with high purities and purification yields, suggesting a successful intein cleavage and a successful releasing of Enzyme α. Although CTC01 chromatography might need some optimization in order to try to avoid the suspected early cleavage induced while introducing the Cleavage Buffer, CTC03 chromatography showed a great result with a coherent activity balance. Since the production of these Intein Samples was performed in 24-well plates, the low volumes of material did also not allow to study the reproducibility of each purification, which might be an important topic for future studies.

After performing the purification studies with the Intein Samples, it was concluded that the most promising fusion construct to study was CTC03, not only for showing the best results regarding purity and purification yield, but also for allowing the purification of an unmodified native Enzyme α. In fact, after inducing CTC03 cleavage it is expected to elute an unmodified native Enzyme α (confirmed by MS). On the other hand, CTC01, for having a linker between the intein and the POI, is expected to enable the purification of an Enzyme α with a scar from the intein-tag (also confirmed by MS). The purification of a POI with a scar is not the ideal situation, in particular for food industry, since it may need new toxicity testing trials, new registrations, and ultimately this POI variant can be considered artificial and encounter resistance in the market. For these reasons, one could suggest designing the fusion constructs without any linker between the intein and the POI. However, that may also influence the fusions expression as it was observed for CTC02 (note that CTC02 only differs from CTC01 for having an extra his-tag and for lacking a linker between Enzyme α and *Mxe* GyrA intein).

The final purification experiment with Fusion Construct 3 Sample was performed to reproduce the previous successful affinity chromatography with the Intein Sample expressing CTC03. However, the obtained results indicated that intein cleavage did not work. The justification behind the successful purification of the CTC03 Intein Sample and the unsuccessful purification of the Fusion Construct 3 Sample is not completely understood. In principle, these samples should be very similar, only differing on their fermentation protocol: the first was produced in 24-well plates, while the second was produced in shake flasks. The only difference in the chromatography protocol was regarding the usage of column capacity: the first successful experiment was performed using 3-6%, while the second (unsuccessful) used 70%. Although no references were found on the literature to justify this result, it may indicate that the fusion construct response to cleavage might be dependent on its concentration inside the column and, consequently, on the column capacity usage. This would justify the different results obtained after purification of the CTC03 Intein Sample and the Fusion Construct 3 Sample.

To better understand how this system works, it would be recommended to begin with a simpler POI, especially without showing interaction between monomers. Enzyme α is a quite complex molecule regarding its quaternary structure, since it is known to show activity both as a dimer and as a tetramer. For that reason, it is difficult to understand whether this POI behaves as a tetramer or as a dimer, and the most reasonable explanation relies on an equilibrium between these two structures. This dimer-

tetramer equilibrium raises several questions: the assumption that the prematurely cleaved fusions in the initial sample cannot be purified may not be that obvious since, in principle, there might exist hybrid dimers or hybrid tetramers formed by both tagged (not cleaved) and untagged (cleaved) fusion monomers. If this is the case, then it complicates the activity balance and the calculation of the purification yields, since not only the intact fusion constructs, but also some unknown amount of native Enzyme α can attach to the resin and be subsequently purified. This theory of hybrid dimers/tetramers is supported by two different obtained results: firstly, the fact that in both CTC01 Intein Sample and in Fusion Construct 3 (CTC03) Sample chromatography not all the expected native Enzyme α was washed out during sample application and washing phases; and secondly, the observation that, during the binding capacity test, the first step showed an activity lower than 50% of the initial sample. In fact, these two results suggested some interaction between the native Enzyme α and the stationary phase, which could be explained by the presence of these hybrid dimers/tetramers.

The last off-column cleavage test showed that only CTC01 did undergo a clear cleavage reaction. The fact that both the Intein Sample expressing CTC02 and the Fusion Construct 3 (CTC03) Sample did not show a proper cleavage, is in accordance with the unsuccessful purifications for both constructs. However, the observation that the Intein Sample expressing CTC03 did not cleave in this batch test is quite contradictory with the purification results achieved with this sample. This contradiction could be explained by a different response from the fusion constructs to cleavage when they are bound to a chitin resin (on-column cleavage) and when they are in solution (off-column cleavage). For these reasons, prior to proceeding to a on-column purification, it would be suggested to study the best conditions for an off-column intein cleavage. For *Mxe* GyrA mini-intein, not only different DTT concentrations but also the use of other nucleophiles (β-ME, cysteine or hydroxylamine) could be studied. For *Mtu* RecA mini-intein, different ranges of pH could also be tested, especially a lower pH (5.5 to 6.0) to induce intein cleavage. Additionally, both mini-inteins are shown to require long incubation periods for cleavage induction, which could also be optimized in combination with an incubation at higher temperature.

Regarding the scalability of the present system, affinity resins are generally expensive and, consequently, are not useful for large-scale purification. Nevertheless, tags, such as the CBD, use as purification matrices polysaccharides that are abundantly present in nature, which could allow the development of low-cost systems for recombinant protein purification [41]. The main concern should be the chitin resin that is not commercially available for large-scale purifications and which can only be regenerated up to five times [38]. However, the coupling of a proper ligand to a pre-activated matrix could be further studied in order to understand the feasibility of such resins for large-scale purification. Another possibility would be to test different tags, such as cellulose-binding or starch-binding domains.

As the last recommendation, one might also consider putting some effort on precipitation tags for non-chromatographic purification. These alternative purifications might be of significant interest since chromatography is typically the most expensive step in protein purification [45]. In fact, CTC06 of the present project included a calcium-responsive precipitation tag. However, it was not further studied since the construct showed a very poor expression.

6. REFERENCES

[1] A. Narváez and E. Domínguez, “Enzymes: Overview,” in *Encyclopedia of Analytical Science*, 2005, pp. 508-522.

[2] P. Panesar, S. Marwaha and H. Chopra, Enzymes in Food Processing: Fundamentals and Potential Applications, New Delhi: International Publishing House, 2010.

[3] W. Aehle, Enzymes in Industry - Production and Applications, 3rd ed., Weinheim: Wiley-VCH, 2007.

[4] B. Sarrouh, T. Marcos Santos, A. Miyoshi, R. Dias and V. Azevedo, “Up-To-Date Insight on Industrial Enzymes Applications and Global Market,” *Journal of Bioprocessing & Biotechniques,* vol. s4:002, 2012.

[5] P. Fernandes, “Enzymes in Food Processing: A Condensed Overview on Strategies for Better Biocatalysts,” *Enzyme Research,* pp. 1-19, 2010.

[6] O. Kirk, T. Borchert and C. Fuglsang, “Industrial enzyme applications,” *Current Opinion in Biotechnology,* vol. 13, pp. 345-351, 2002.

[7] BBC Research, “Global Markets,” *Global Markets for Enzymes in Industrial Applications,* 2014.

[8] BBC Research, “Global Markets,” *Global Markets for Enzymes in Industrial Applications,* 2018.

[9] Z. Berk, “Reaction Kinetics,” in *Food Processing Engineering and Technology*, Academic Press, 2009, pp. 115-127.

[10] N. Saraswathy and P. Ramalingam, “Introduction to Proteomics: Concepts and Techniques,” in *Genomics and Proteomics*, 2011, pp. 147-158.

[11] M. Brasher and R. Thorpe, “Isoelectric Focusing,” in *Encyclopedia of Immunology*, 1998, pp. 1510-1514.

[12] B. Ersson, L. Rydén and J. Janson, “Introduction to Protein Purification,” in *Protein Purification: Principles, High Resolution Methods, and Applications*, Hoboken, N. J., John Wiley & Sons, 2019.

[13] R. Freitag, “Chromatographic Techniques in the Downstream Processing of Proteins in Biotechnology,” in *Animal Cell Biotechnology - Methods and Protocols*, Humana Press, 2014, pp. 419-458.

[14] M. Kalyanpur, “Downstream Processing in the Biotechnology Industry,” *Molecular Biotechnology,* vol. 22, no. 1, pp. 87-98, 2002.

[15] O. Coskun, “Separation Techniques: Chromatography,” *Northern Clinics of Istanbul,* vol. 3, no. 2, pp. 156-160, 2016.

[16] J. Janson and J. Jönsson, “Introduction to Chromatography,” in *Protein Purification: Principles, High Resolution Methods, and Applications*, Hoboken, N.J., John Wiley & Sons, 2019.

[17] GE Healthcare, “Ion Exchange Chromatography - Principles and Methods,” GE Healthcare, 2016.

[18] GE Healthcare, “Affinity Chromatography - Principles and Methods,” GE Healthcare, 2007.

[19] J. Bornhorst and J. Falke, “Purification of Protein using Polyhistidine Affinity Tags,” in *Methods in Enzymology*, vol. 326, 2000, pp. 245-254.

[20] F. Batista-Vieira, J. Janson and J. Carlsson, “Affinity Chromatography,” in *Protein Purification: Principles, High Resolution Methods, and Applications*, Hoboken, N. J., John Wiley & Sons, 2019.

[21] S. Hofer, A. Ronacher, J. Horak, H. Graalfs and W. Lindner, “Static and dynamic binding capacities of human immunoglobulin G on polymethacrylate based mixed-modal, thiophilic and hydrophobic cation exchangers,” *Journal of Chromatography A,* vol. 1218, no. 49, pp. 8925-8936, 2011.

[22] M. Carlsson, “How to interpret protein binding capacity: chromatography resins,” GE Healthcare Life Sciences, 2018. [Online]. Available: https://www.gelifesciences.com/en/us/news-center/protein-binding-chromatography-resins-10001. [Accessed 16 08 2019].

[23] GE Healthcare Life Sciences, “How to determine dynamic binding capacity (DBC) of chromatography resins,” GE Healthcare Life Sciences, 2018. [Online]. Available: https://www.gelifesciences.com/en/nz/news-center/how-to-determine-dynamic-binding-capacity-10001. [Accessed 16 08 2019].

[24] E. Hosseini, R. Moniri, Y. Goli and H. Kashani, “Purification of Antibacterial CHAPK Protein Using a Self-Cleaving Fusion Tag and Its Activity Against Methicillin-Resistant Staphylococcus aureus,” *Probiotics and Antimicrobial Proteins,* vol. 8, pp. 202-210, 2016.

[25] A. Telenti, M. Southworth, F. Alcaide, S. Daugelat, W. Jacobs and F. Perler, “The Mycobacterium xenopi GyrA protein splicing element: characterization of a minimal intein,” *Journal of Bacteriology,* vol. 179, no. 20, pp. 6378-6382, 1997.

[26] S. Chong, F. Mersha, D. Comb, M. Scott, D. Landry, L. Vence, F. Perler, J. Benner , R. Kucera, C. Hirvonen, J. Pelletier, H. Paulus and M. Xu, “Single-column purification of free recombinant proteins using a self-cleaving affinity tag derived from a protein splicing element,” *Gene,* vol. 192, pp. 271-281, 1997.

[27] Z. Du, J. Liu, C. Albracht, A. Hsu, W. Chen, M. Marieni, K. Colelli, J. Williams, J. Reitter, K. Mills and C. Wang, “Structural and Mutational Studies of a Hyperthermophilic Intein from DNA Polymerase II of Pyrococcus abyssi,” *Journal of Biological Chemistry,* vol. 286, no. 44, pp. 38638-38648, 2011.

[28] V. Derbyshire, D. Wood, W. Wu, J. Dansereau, J. Dalgaard and M. Belfort, “Genetic definition of a protein-splicing domain: Functional mini-inteins support structure predictions and a model for intein evolution,” *Proceedings of the National Academy of Sciences,* vol. 94, pp. 11466-11471, 1997.

[29] D. Wood, W. Wu, G. Belfort, V. Derbyshire and M. Belfort, “A genetic system yields self-cleaving inteins for bioseparations,” *Nature Biotechnology,* vol. 17, pp. 889-892, 1999.

[30] T. Evans, J. Benner and M. Xu, “Semisynthesis of cytotoxic proteins using a modified protein splicing element,” *Protein Science,* vol. 7, pp. 2256-2264, 1998.

[31] K. Mills and H. Paulus, “Biochemical Mechanisms of Intein-Mediated Protein Splicing,” in *Nucleic Acids and Molecular Biology*, Heidelberg, Springer, 2005, pp. 233-255.

[32] S. Frutos, M. Goger, B. Giovani, D. Cowburn and T. Muir, “Branched intermediate formation stimulates peptide bond cleavage in protein splicing,” *Nature Chemical Biology,* vol. 6, no. 7, pp. 527-533, 2010.

[33] Z. Du, Y. Zheng, M. Patterson, Y. Liu and C. Wang, “pKa Coupling at the Intein Active Site: Implications for the Coordination Mechanism of Protein Splicing with a Conserved Aspartate,” *Journal of the American Chemical Society,* vol. 133, pp. 10275-10282, 2011.

[34] M. Coolbaugh, M. Shakalli Tang and D. Wood, “High-throughput purification of recombinant proteins using self-cleaving intein tags,” *Analytical Biochemistry,* vol. 516, pp. 65-74, 2017.

[35] M. Southworth, K. Amaya, T. Evans, M. Xu and F. Perler, “Purification of Proteins Fused to Either the Amino or Carboxy Terminus of the Mycobacterium xenopi Gyrase A Intein,” *BioTechniques,* vol. 27, no. 1, pp. 110-120, 1999.

[36] S. Chong, G. Montello, A. Zhang, E. Cantor, W. Liao, M. Xu and J. Benner, “Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step,” *Nucleic Acids Research,* vol. 26, no. 22, pp. 5109-5115, 1998.

[37] M. Coolbaugh and D. Wood, “Purification of *E. coli* Proteins Using a Self-Cleaving Chitin-Binding Affinity Tag,” in *Protein Affinity Tags: Methods and Protocols*, vol. 1177, New York, Humana Press, 2014, pp. 47-58.

[38] New England Biolabs, “IMPACT™ Kit Instruction Manual. Protein Expression & Analysis.,” New England Biolabs, 2017.

[39] R. Bott, M. Ultsch, A. Kossiakoff, T. Graycar, B. Katz and S. Power, “The Three-dimensional Structure of Bacillus amyloliquefaciens Subtilisin at 1.8 A and an Analysis of the Structural Consequences of Peroxide Inactivation,” *The Journal of Biological Chemistry,* vol. 263, no. 16, pp. 7895-7906, 1988.

[40] D. Wood, “Self-Cleaving Affinity Tags and Methods of Use”. Patent WO 2005/086654 A2, 2005.

[41] D. Guillén, S. Moreno-Mendieta, P. Aguilera, S. Sánchez, A. Farres and R. Rodríguez-Sanoja, “The starch-binding domain as a tool for recombinant protein purification,” *Applied Microbiology and Biotechnology,* vol. 97, pp. 4141-4148, 2013.

[42] T. Ikegami, T. Okada, M. Hashimoto, S. Seino, T. Watanabe and M. Shirakawa, “Solution Structure of the Chitin-binding Domain of Bacillus circulans WL-12 Chitinase A1,” *The Journal of Biological Chemistry,* vol. 275, no. 18, pp. 13654-13661, 2000.

[43] K. Terpe, “Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems,” *Applied Microbiology and Biotechnology,* vol. 60, pp. 523-533, 2003.

[44] S. Gort and S. Maloy, “Purification of a hexahistidine-tagged protein using L-histidine as the eluent,” *Technical Tips Online,* vol. 3, pp. 54-55, 1998.

[45] O. Shur, K. Dooley, M. Blenner, M. Baltimore and S. Banta, “A designed, phase changing RTX-based peptide for efficient bioseparations,” *BioTechniques,* vol. 54, no. 4, pp. 197-206, 2013.

[46] GE Healthcare Life Sciences, “Tricorn empty chromatography columns,” GE Healthcare Life Sciences, 2019. [Online]. Available: https://www.gelifesciences.com/en/us/shop/chromatography/columns/empty-columns-for-lab-scale/tricorn-empty-chromatography-columns-p-06216 . [Accessed 25 08 2019].

[47] GE Healthcare, “Instructions 28-9574-96 AD. HiTrap™ TALON® crude, 1 ml and 5 ml TALON Superflow™,” GE Healthcare, 2014.

[48] Sigma-Aldrich, “Biological Buffers Selection Guide,” [Online]. Available: https://www.sigmaaldrich.com/life-science/core-bioreagents/biological-buffers.html#selectionguide. [Accessed 25 08 2019].

[49] N. Good, G. Winget, W. Winter, T. Connolly, S. Izawa and R. Singh, “Hydrogen Ion Buffers for Biological Research,” *Biochemistry,* vol. 5, no. 2, pp. 467-477, 1966.

[50] Thermo Fisher Scientific, “Novex® Pre-Cast Gel Electrophoresis Guide. General information and protocols for using Novex® pre-cast gels,” Invitrogen, 2010.

[51] RCSB PDB, “Rcsb.org,” 2019. [Online]. Available: https://www.rcsb.org/. [Accessed 29 08 2019].

[52] G. Kleiner-Grote, J. Risse and K. Friehs, “Secretion of recombinant proteins from *E. coli*,” *Engineering in Life Sciences,* vol. 18, pp. 532-550, 2018.

[53] M. Shehadul Islam, A. Aryasomayajula and P. Selvaganapathy, “A Review on Macroscale and Microscale Cell Lysis Methods,” *Micromachines,* vol. 8, no. 83, 2017.

[54] J. Geciova, D. Bury and P. Jelen, “Methods for disruption of microbial cells for potential use in the dairy industry - a review,” *International Dairy Journal,* vol. 12, pp. 541-553, 2002.

[55] J. Pierce, C. Turner, E. Keshavarz-Moore and P. Dunnill, “Factors determining more efficient large-scale release of a periplasmic enzyme from *E. coli* using lysozyme,” *Journal of Biotechnology,* vol. 58, pp. 1-11, 1997.

[56] F. Mergulhão, D. Summers and G. Monteiro, “Recombinant protein secretion in Escherichia coli,” *Biotechnology Advances,* vol. 23, pp. 177-202, 2005.

[57] Proteintech, “Lysate Preparation: How Do I Optimize My Extraction?,” Proteintech, 2019. [Online]. Available: https://www.ptglab.com/news/blog/lysate-preparation-how-do-i-optimize-my-extraction/. [Accessed 15 09 2019].

[58] Thermo Fisher Scientific, “Traditional Methods of Cell Lysis,” Thermo Fisher Scientific, 2019. [Online]. Available: https://www.thermofisher.com/nl/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/traditional-methods-cell-lysis.html. [Accessed 15 09 2019].

[59] A. Gillies, J. Hsii, S. Oak and D. Wood, “Rapid Cloning and Purification of Proteins: Gateway Vectors for Protein Purification by Self-Cleaving Tags,” *Biotechnology and Bioengineering,* vol. 101, pp. 229-240, 2008.

[60] W. Wu, K. Miller, M. Coolbaugh and D. Wood, “Intein-mediated one-step purification of Escherichia coli secreted human antibody fragments,” *Protein Expression and Purification,* vol. 76, pp. 221-228, 2011.

[61] D. Wood, V. Derbyshire, W. Wu, M. Chartrain, M. Belfort and G. Belfort, “Optimized Single-Step Affinity Purification with a Self-Cleaving Intein Applied to Human Acidic Fibroblast Growth Factor,” *Biotechnology Progress,* vol. 16, pp. 1055-1063, 2000.

[62] W. Wu, A. Gillies, J. Hsii, L. Contreras, S. Oak, M. Perl and D. Wood, “Self-cleaving Purification Tags Re-engineered for Rapid Topo® Cloning,” *Biotechnology Progress,* vol. 26, no. 5, pp. 1205-1212, 2010.