IPS Instituto Politécnico de Setúbal
Escola Superior de Tecnologia do Barreiro

CATARINA ALEXANDRA OLIVEIRA DOMINGOS

PHOTOBIOLOGY OF LIGHT-SENSITIVE TRANSCRIPTION FACTORS

Relatório de estágio do Mestrado em Engenharia Biológica e Química

SUPERVISOR
Doutor Gustavo Fuertes Vives

CO-SUPERVISOR
Professor Doutor Bohdan Schneider

COORIENTADORA
Professora Doutora Marta Campos Justino

Dezembro 2019

CATARINA ALEXANDRA OLIVEIRA DOMINGOS

PHOTOBIOLOGY OF LIGHT-SENSITIVE TRANSCRIPTION FACTORS

JÚRI

Presidente: Professora Doutora Maria de Lurdes de Figueiredo Gameiro (ESTBarreiro/IPS)
Coorientadora: Professora Doutora Marta Campos Justino (ESTBarreiro/IPS)
Vogal: Doutora Célia Valente Romão (ITQB, Universidade Nova de Lisboa)

Dezembro 2019

CATARINA ALEXANDRA OLIVEIRA DOMINGOS

PHOTOBIOLOGY OF LIGHT-SENSITIVE TRANSCRIPTION FACTORS

Internship Report of the Master's in Biological and Chemical Engineering

SUPERVISOR
Doctor Gustavo Fuertes Vives

CO-SUPERVISOR
Professor Bohdan Schneider

CO-ADVISOR
Professor Marta Campos Justino

December 2019

To my Brother

Important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.

Sir William Bragg

ACKNOWLEDGMENTS

At the end of my Master's thesis, I have to thank all the people who have contributed to make this work satisfactorily complete.

Firstly I have to thank the BIOCEV and IBT research centre for welcoming me over the 5 months of my Master's internship, as well as the head of the Structural Biology and Protein Engineering Program and leader of the Intermolecular Recognition of Proteins and Nucleic Acids Group , the Doctor Engineer Bohdan Schneider who accepted me to do the internship in one of his teams giving me this great opportunity.

Next, I have to thank Gustavo Fuertes Vives, the program leader of the "Dynamics of Biological Processes" project and my advisor at the research center. Thanks for all the support and guidance he has given me over the 5 months of my internship, as well as all the patience and time he spent to teach me, always helping me to overcome my difficulties, to correct my mistakes and to help me not committing them again, by the freedom and willingness to elaborate the experiments and to discuss the most diverse subjects, as well as, for all the support.

I also have to thank the whole team that welcomed me, paying special attention to my teammate and student of education Aditya Suresh Chaudhari for all the help and support he gave me throughout my days in the lab, for all the patience and availability to help me with all the tasks.

Next, I have to thank my college advisor Professor Marta Justino for all the support and guidance in developing the thesis.

Finally, I have to thank my family and friends for all the support and strength given in these months of work abroad. Especially to my younger brother Salvador and my sister Beatriz, my parents Gina and Rui and my grandmother Maria da Conceição and my cousin Inês, who despite the distance gave me strength to finish this mission. And thanks to my friend Barbara Albuquerque who embarked on this adventure with me. And finally, to my friends Barbara Silva, Catarina Dias, Vanessa Cova, Adrielly Carreira and David Sousa for all their support.

ABSTRACT

Protein structure and function is a major research area in life sciences. A novel field emerged since the discovery of proteins that are photosensitive. These photoreceptor proteins undergo light-induced structural changes, performing the most varied tasks and contributing to the operation of a large number of biological processes.

The importance of this type of proteins is because in the near future it will possible to use them in different contexts and in a way that is beneficial to humans.

The final curricular internship of the Master in Biological and Chemical Engineering was held at the Institute of Biotechnology of the BIOCEV research center located in Vestec, Czech Republic. I was received for a period of 5 months and was part of a multicultural team working on a project entitled “Dynamics of Biological Processes”.

Throughout these 5 months of Master’s degree internship, the main objective was to conduct a deeper study on the photoinduced changes of CarH and EL222 proteins. For both proteins, it is important to discover the conformational changes that occur between the dark state and the light state

A realistic work plan was developed that is appropriate to the capabilities of both the team and the research center. At the end of the internship period, significant results were obtained . In particular, I demonstrated the possibility of introducing non-canonical amino acids in EL222 with sufficient yield for future infrared spectroscopy experiments. I also showed the feasibility of studying CarH structural dynamics by infrared spectroscopy.

KEYWORDS: Protein structure and dynamics, photoreceptors, protein-DNA interactions, non-canonical amino acids, infrared spectroscopy.

GENERAL INDEX

INDEX OF FIGURES

INDEX OF TABLES

SYMBOLS AND ABBREVIATIONS

AdoCbl	5'-adenosylcobalamin
CD	Circular dichroism
CNF	4-cyano-L-phenylalanine
FMN	Flavin mononucleotide
FTIR	Fourier-transform infrared
HTH	Helix-turn-helix
K_D	Dissociation constant
LOV	Light-oxygen-voltage domain
MS	Mass spectrometry
MST	Microscale thermophoresis
PCR	Polymerase chain reaction
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
λ	Wavelength

1. BIOCEV AND IBT

The BIOCEV, Biotechnology and Biomedicine Centre of the Academy of Sciences and Charles University in Vestec (Czech Republic), is a scientific centre created with the collaboration of six Institutes of the Czech Academy and Sciences (CAS) and two faculties of Charles University in Prague. In January 2008 the BIOCEV project started and there are currently 5 research programmes with 440 researchers, students and technical staff employed. This project has a budget of 2,3 billion of CZK, which in euros round about 897,000,000.00 € [1].

BIOCEV has three guide pillars [1]:

- Teaching and education
- Research and development
- Transfer of research results into practice

The aim of the Institute of Biotechnology (IBT) of the CAS is basic research in the field of molecular biological sciences at the top level and prospectively the transfer of biotechnological methods and molecular tools for the diagnosis and treatment of the pathological state of the cell into human medicine or other important areas of human activity [2].

At this investigation centre I developed my curricular internship for the Master Degree in the department of Structural Biology and Protein Engineering in the team of “Dynamics of Biological processes”. This internship was done under the Erasmus program. This project aims to understand how light controls the structure, dynamics and activity of photosensitive proteins. The EL222 and CarH proteins were studied for this purpose.

In table 1 it is possible to observe the evolution of the project during the 5 months of curricular internship and all the steps that were taken.

Table 1 - Evolution of the curricular internship.

Internship weeks																				
Steps	March				April				May				June				July			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Cloning																				
Mutagenesis																				
Protein expression																				
Protein purification																				
Radiation Study																				
UV-Vis																				
MST																				
FTIR																				
Retreat																				
CD																				

2. INTRODUCTION

2.1. Photoreceptors and photosensitivity

Over time, it has been found that a large group of living organisms are able to respond to stimuli caused by light. This is in part due to the existence of a new class of light-sensitive photoreceptor proteins.

It is known that excessive irradiation can be harmful for life because it can degrade DNA, proteins and other cellular structures. In order to combat this, by using photoreceptors, cells develop a mechanism for the synthesis of protective pigments or even the direct repair of DNA damaged by light [3-4].

There are ten families of photoreceptors, these being rhodopsins, phytochromes, photoactive yellow proteins, light-oxygen-voltage (LOV) proteins, blue-light sensors using flavin (BLUF), cryptochromes and the recently discovered UVR8, B_{12}-dependent photoreceptors, orange carotenoid protein (OCP) and UV receptor LITE-1 [3-4].

Each of these families is associated with a type of chromophore, which is responsible not only for the colour but also for the conformational changes that the molecule suffers in contact with light. For instance, for the rhodopsins family we have retinal; for phytochromes it is bilin; for LOV, criptochromes and BLUF it is flavins; and finaly for photoactive yellow proteins it is p-coumaric acid (pCA) [3-4].

In photoreceptors, the formation of the signalling states is typically caused by proton and/or electron transfer (electrostatics), changes in hydrogen bonding, or the distortion of the chromophore and strain/steric clashes between the chromophore and groups in its binding pocket [3-4].

2.2. LOV Domain

The LOV domain belongs to a superfamily of the Per-ARNT-Sim domain (PAS) domain. The LOV domain (Figure 1A) presents biological activity in the presence of blue light, being used as a response for phototropism, gene regulation and stress response. In the initial dark state, the LOV domain is associated with flavin mononucleotide (FMN) through well-conserved hydrogen bonds usually involving using two residues of asparagine and two residues of glutamine, as can be seen in Figure 1B [3,5].

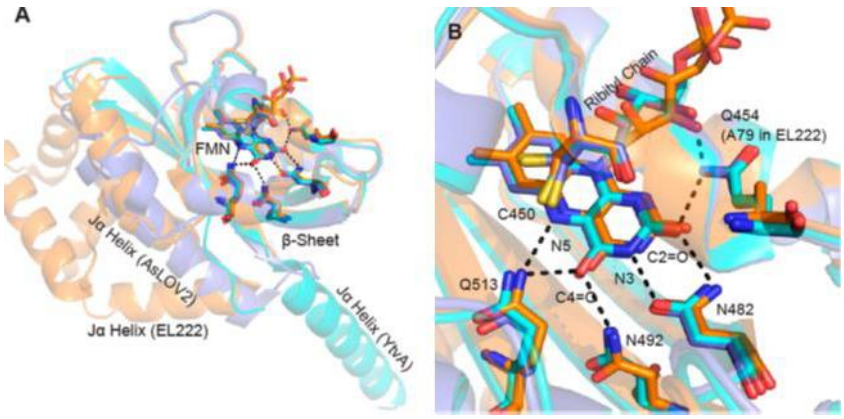


Figure 1 - LOV domain structure [5].

When the LOV domain is exposed to blue light, it undergoes some transformations. After illumination, occurs the formation of a covalent adduct between C4a atom of FMN and the sulphur atom of a conserved cysteine, and the simultaneous protonation of the N5 of FMN. After such first change, a rotation of a glutamine residue is expected to occur, causing changes in a β-sheet of the LOV domain. All these changes lead to an activation of the LOV domain causing the biological activity to begin. The cysteine present in the structural transformations of the LOV domain is not essential for flavin binding, but it is required for the reversible reaction upon light stimulation. The dark recovery of the domain depends on pH, salt concentration or addition of low molecular weight bases. Figure 2 shows the possible structural transformations of the LOV domain upon light irradiation [3,5].

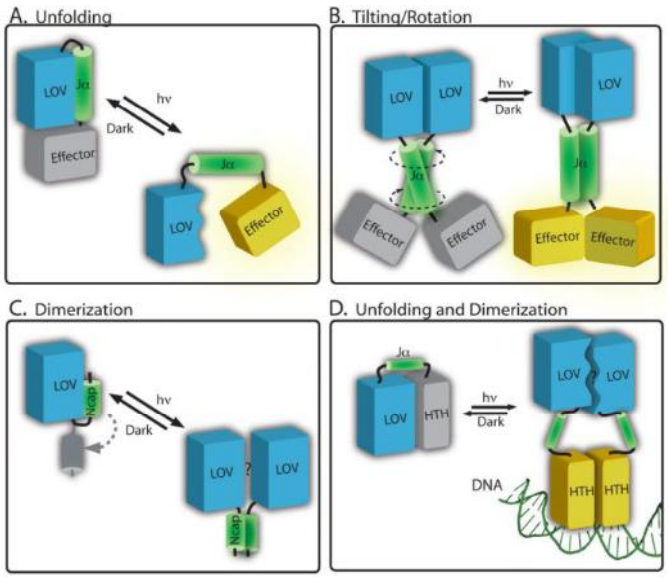


Figure 2 – Possible molecular mechanisms of LOV domain signalling. (A) Phototropine-type signalling. (B) Rotation of the effector. (C) Rearrangement. (D) A combination of mechanisms A and C [3].

The LOV domain was initially discovered in plant phototropins but later also in some fungi and algae and more recently in bacteria. In the case of bacteria, the LOV domain is coupled to other "effector" domains responsible for the signalling output used in the stress response. For instance, the LOV domain found in EL222 protein is fused to a helix-turn-helix (HTH) DNA-binding domain [3].

2.3. B_{12}-dependent domains

One of the 10 families of photoreceptors is B_{12}-dependent and its prototype, CarH, controls light-dependent gene expression [6]. CarH contains two domains: a B_{12}-binding light-sensitive domain and a HTH DNA-binding domain. The B_{12}-based CarH domain is sensitive to UV, blue and green light but not red light [7-8].

The CarH protein has as chromophore 5'-adenosylcobalamin (AdoCbl) or coenzyme B_{12}, which is one of the active forms of vitamin B_{12}. The cobalt (Co) atom of AdoCbl is hexa-coordinated by four in-plane nitrogen atoms in the coring ring, the Ado moiety as the upper axial ligand and dimethylbenzimidazole as the lower axial ligand. The extremely weak covalent bond between the Co and the Ado moiety is the basis of AdoCbl photochemistry [9]. Photoreceptors are in the cytoplasm and once the chromophore is excited with light, the proteins changes to its "light" state [7-8].

As can be seen in Figure 3, while kept in the dark, the CarH protein binds non-covalently to AdoCbl and associates with DNA thereby preventing RNA polymerase (RNAP) binding and blocking gene transcription. Upon exposure to light, CarH is no longer able to bind the DNA strand leaving the path free for RNAP binding to occur and consequent transcription of the gene. The light-induced reactions are essentially irreversible [7-8].

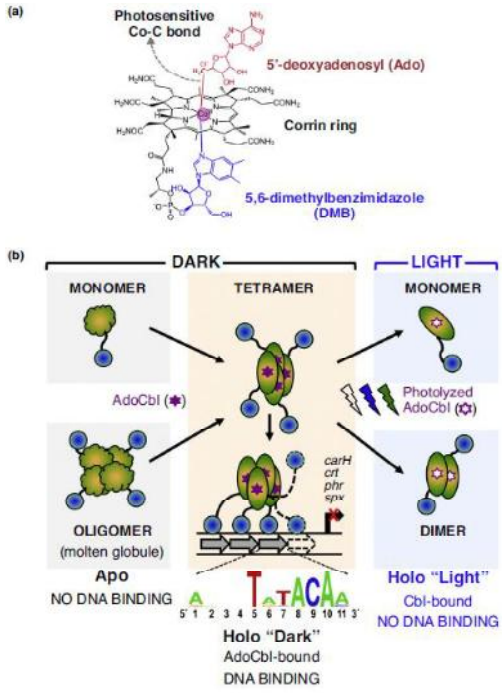


Figure 3 - Mechanism of action of B_{12}-dependent photoreceptors. (a) Chemical structure of AdoCbl bound to CarH protein in the dark (b) Effect of light on the oligomerization state of CarH and its interaction with DNA [7].

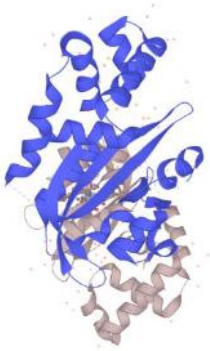
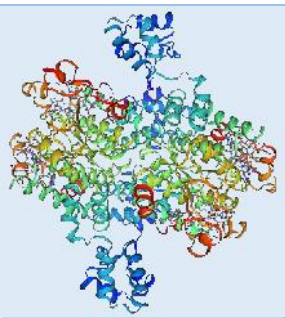
The CarH protein is responsible for regulating the expression of its own gene and the synthesis of carotenoids to protect the cells against photooxidative damage. It is also capable of regulating the expression of photolyase in order to repair UV-induced pyrimidine dimers in DNA or other general redox response [7-8].

Due to its recent discovery there is still much to discover about this mechanism, however in the future it might be possible to use proteins such as CarH to control the expression of target genes [7-8].

2.4. Proteins under study

In order to study these light-responsive DNA-binding photoreceptors, a set of proteins was selected in order to represent each of the domains. For each protein, a set of tests was performed in order to characterize its structure and mechanism. Table 2 contains a summary of basic information on the studied proteins.

Table 2 - Set of proteins under study [10-11].

Target photoreceptor proteins						
Name of the Protein	Organisms	Function	Light-sensitive domain	Ligand	Activity regulation	Dark-state structure
EL222	*Erythrobacter litoralis* (Bacteria)	Transcriptional activator	LOV	FMN	Dimerization DNA binding	
CarH	*Thermus thermophilus* (Bacteria)	Transcriptional Repressor	Cobalamin (vitamin B_{12}) binding domain	AdoCbl	Tetramer dissociation DNA unbinding	

EL222 protein also has its own mechanism (Figure 4). In its dark state, the protein is monomeric and does not bind to DNA [12]. Therefore, there is no transcription. Once illuminated with blue light, EL222 changes from a monomer to a dimer [13-14]. In this state, the protein is capable of binding to DNA thus activating transcription [15-16]. In this case, the process is reversible, and the protein is expected to return to the dark state.

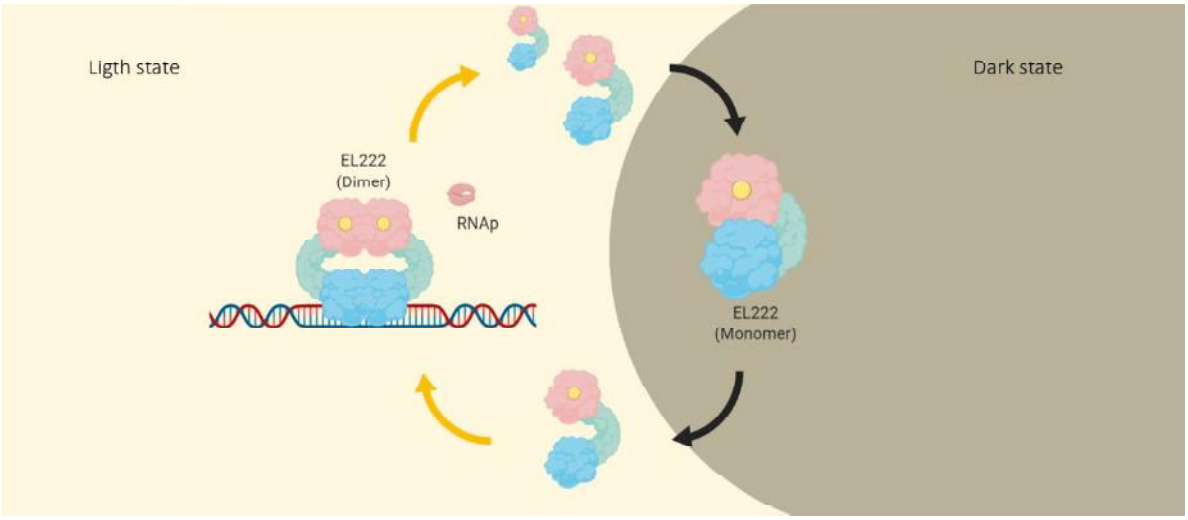


Figure 4 - Mechanism of EL222 protein. Creating using https://app.biorender.com/signin [17].

CarH protein has a unique mechanism (Figure 5). In the dark state, it has a tetrameric AdoCbl-bound form and has high affinity for DNA binding thus preventing transcription. However, when it gets light and changes to its light state, it dissociates into monomers and/or dimers, loosing its ability to bind DNA and thus RNAP can transcribe genes. Such process is not spontaneously reversible in the dark.

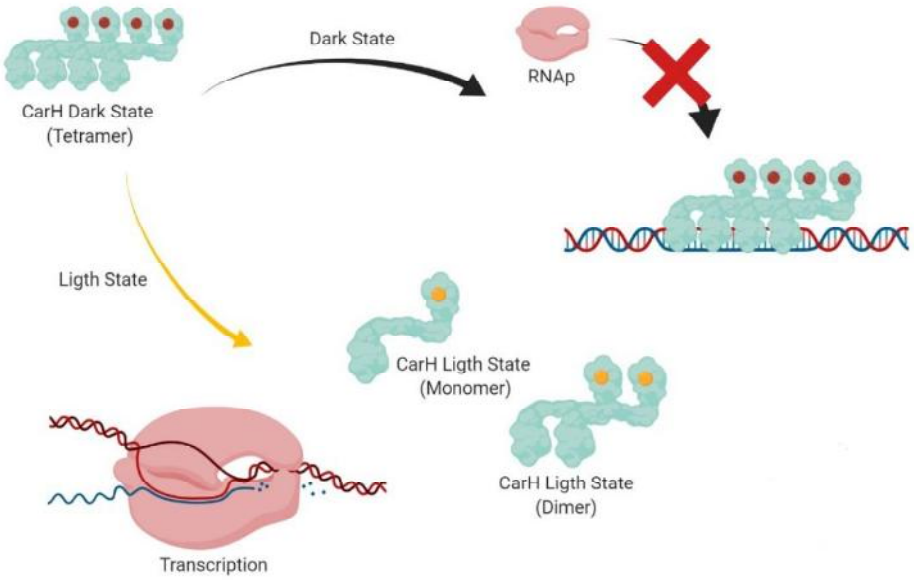


Figure 5 - Mechanism of CarH protein Creating using https://app.biorender.com/signin [17].

2.5. Protein Studies

2.5.1. Genetic code expansion (GCE).

GCE technology was used in order to introduce infrared reporters at any desired position within the sequence with the goal to obtain site-specific structural information by FTIR spectroscopy (see below). The technology is based on an engineered aminoacyl tRNA synthetase/tRNA pair that is able to recognize a UAG codon in the messenger RNA and introduce a non-canonical amino acid (an amino acid other than any of the 20 canonical ones) in the growing protein chain [18]. In such a way, the recoded UAG codon does not mean "stop translation" anymore but rather "incorporate a non-canonical amino acid". GCE was used to introduce the non-canonical amino acid 4-cyano-phenylalanine (CNF) that carries the infrared-sensitive nitrile (C≡N) probe [19].

2.5.2. Mass spectrometry (MS)

MS (Figure 6) is a widely used technique for protein identification because it is a fast process that uses little protein amounts (5mg/ml in 20 µL) and its results are very reliable.

To carry out this type of experiment, a protein sample is prepared as pure as possible in its usual buffer, this sample will contain the protein, ligands and some contaminants. MS determines the mass / charge ratio (m/z) of the intact protein. At the end of the process, a graph is presented and analysed using statistical software in order to obtain the list of proteins present in the sample. Because of the relatively harsh experimental conditions, non-covalently bound cofactors (e.g. FMN) are typically not observed. In other words, the observed masses tend to reflect the masses of the polypeptide chain only (apo-protein) [20-21].

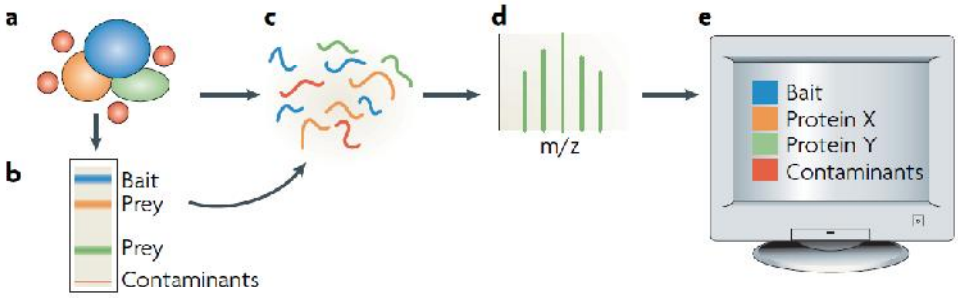


Figure 6 - Scheme of the mass spectrometry experiment. A) sample; B) SDS-PAGE for confirmation of the present compounds; C) proteolysis; D) The first scan; E) Final result [20].

2.5.3. Fourier-transform infrared (FTIR) Spectroscopy

FTIR spectroscopy is a technique used to determine protein binding, secondary structure and conformational changes.

This technique analyses the vibrations of virtually all chemical bonds present in the sample. In particular, the so-called Amide I band provides information on protein conformation and the possible changes that occur upon ligand binding or light irradiation [22-23].

The first step to perform the experiment is to record the reference spectrum of the buffer in a suitable cell. Then a sample is placed in the same cell, and a measurement is made (sample spectrum minus buffer spectrum) with a defined number of scans. The higher the number of scans, the higher the signal-to-noise ratio of the final spectrum. When protein conformational changes are to be studied, the protein is brought into contact with the change factor (e.g. light) and rescanned. Comparing the two data sets i.e. light-state spectrum minus dark-state spectrum, gives the changes that the protein undergoes [22].

However, in addition to the aforementioned factors, this technique requires the use of a high protein concentration (typically in the millimolar range, and 15 to 30 µL for each measurement) and possible buffer switching to deuterated water (D_2O or "heavy" water) instead of "normal" water (H_2O) to eliminate water interference (this step causes a small loss of protein) [22].

2.5.4. Microscale Thermophoresis (MST)

MST is a technique used for the quantitative determination of interactions between proteins, protein and DNA, and DNA hybridization. For this experiment, a few microliters of protein at nanomolar concentration in a suitable aqueous buffer is sufficient. The samples are introduced in a capillary. Because the assay is based on fluorescence detection, one of the two interaction partners must be fluorescent. In our case, we used a DNA fragment labelled with the Cy5 dye [24-25].

This MST experiment itself consists in focusing an infrared beam of light into a specific point of the capillary thus creating a temperature gradient. This temperature gradient causes a change in the sample the magnitude of which depends on the presence or absence of biomolecular interactions [24-25].

Once the capillary is placed in the machine, a first measurement is made to determine the initial fluorescence since at this point the sample is homogeneous. Once the laser is switched on, the T-jump takes place where the molecules will move due to the temperature gradient. After the laser is switched off, a second measurement is taken as the sample returns to the initial state. Using the obtained values and appropriate mathematical formulas, the binding affinity between protein and DNA can be obtained [24-25].

The main advantages of using this technique are the low amount of sample required and the short waiting times.

2.5.5. Circular dichroism (CD) spectroscopy.

CD experiment is a rapid test, which uses little protein (20 µL of protein for each measurement) at not very high concentrations (not below 100nM) and provides information about the secondary structure of the protein under study.

This experiment is based on the circular dichroism phenomena, that is to say, the unequal absorption of right and left circularly polarized light by proteins [26].

To perform this experiment, it is sufficient to use 20 µl of protein sample at a concentration between 0.005 and 5 mg/mL having a purity of at least 95%. The protein may be in an aqueous buffer as it does not interfere with this experiment [26].

3. MATERIALS AND METHODS

3.1. Materials, reagents and equipaments

For working in the laboratory and performing the methods described in section 3.2., a set of equipment and reagents was necessary.

All reagents used in this work were from sigma-aldrich, PENTA, New England BioLabs, Thermo Scientific, TaKaRa and Merck. The brand used depended of the availability of the reagent.

For the equipment, it was necessary to use instruments like centrifuges, balances, shakers, incubators, PCR thermocycler machine, gel filtration chromatography system, optical devices, among others, were mainly from Thermo Scientific and ThorLabs. Figures in the Annex illustrate these equipment.

3.2. Biological materials

For the production of the proteins under study it was necessary to use DNA plasmids (see Annex) and bacteria. Table 3 summarizes the cells used and their main applications.

Table 3 - Biological material used for the production of the proteins under study [27-28].

Strain	Aplication
E.coli BL21(DE3) **New England BioLabs®**	Protein expression
E. coli electrocompetent cells clooni **New England BioLabs®**	Cloning Plasmid maintenance

3.3. Methods

For the production and purification of a proteins, whether wild-type or with mutations, it is necessary to follow a set of protocols to obtain acceptable quality and quantity for subsequent analytical techniques.

Some of these protocols will be described in a simplified standard form in order to be applicable to a larger number of proteins, however it must be kept in mind that they were always adapted to the particular protein under study.

3.3.1. Site-directed mutagenesis

Mutagenesis is a technique used to make alterations in the DNA sequence that will in turn alter the sequence (primary structure) of the encoded protein. In the specific case of EL222 protein, this method was applied in order to introduce amber stop codons (TAG) and subsequent incorporation of non-canonical amino acids in chosen parts of the protein by GCE technology.

First, it was necessary to run a reaction to introduce the mutation in the plasmid. The reaction mixtures were prepared as described in Table 4, to a final volume of 50 μL.

Table 4 - Reaction mix for mutagenesis.

Mutagenesis Mix	
5x Reaction Buffer	10 μL
Plasmid DNA Template (10ng)	~ 0.2-1 μL
Primer #1 (forward)	2.5 μL
Primer #2 (reverse)	2.5 μL
dNTP Mix	1 μL
Water	~ 33 μL
DNA Polymerase	0.5 μL

After the PCR, 1 μL of DpnI restriction enzyme (New England BioLabs®) is added and incubated for 3 hours at 37°C. In this step, the methylated plasmid is digested and the new plasmid with the desired mutation from the reaction stays intact. This step is important because only the template non-mutated plasmid produced in bacteria is methylated and for this reason it is possible to remove it from the mixture. Next, the reaction mix is incubated for 20 minutes at 80°C to inactivate the restriction enzyme. Finally, 50μL of competent cells are transformed by electroporation (MicroPulser Electroporator BIO-RAD®) with 1 μL of mixture. In Figure 7 the mutagenesis protocol is schematically shown.

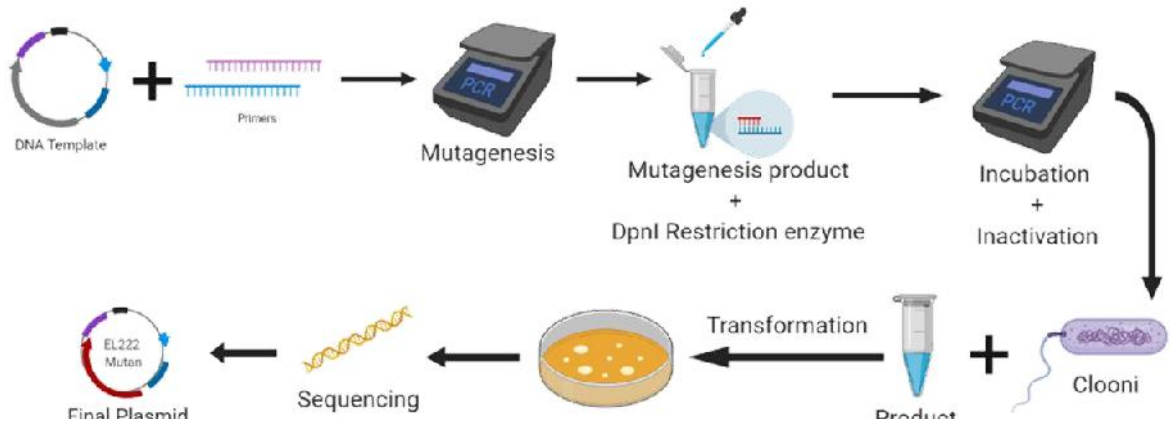


Figure 7 - Protocol for performing mutagenesis. First, select the appropriate template DNA (plasmid), then use the PCR machine to generate DNA copies containing the desired mutation, then apply the restriction enzyme to remove the template plasmid and then inactivate the enzyme. Finally, transformation was performed using competent cells (Clooni) and the final plasmid was verified by DNA sequencing and stored. Creating using https://app.biorender.com/signin [17].

3.3.2. Cloning

The first step in producing a protein is to have the necessary genetic information in an expression plasmid that it is then inserted into a bacteria suitable for expression (*E. coli BL21(DE3)*).

To this end, the cloning protocol described below applies where the gene containing the information necessary for protein production is inserted into a new vector. This protocol applies to EL222 protein only. The plasmids encoding CarH wild type protein (#030, see Annex and CarH cobalamin domain (#059, see Annex) were a kind gift of Dr. Subramanian Padmanabhan Iyer (Instituto de Quimica-Fisica Rocasolano, CSIC, Madrid, Spain). The cloning protocol is performed on two to three consecutive days. Figure 8 shows schematically the process for cloning.

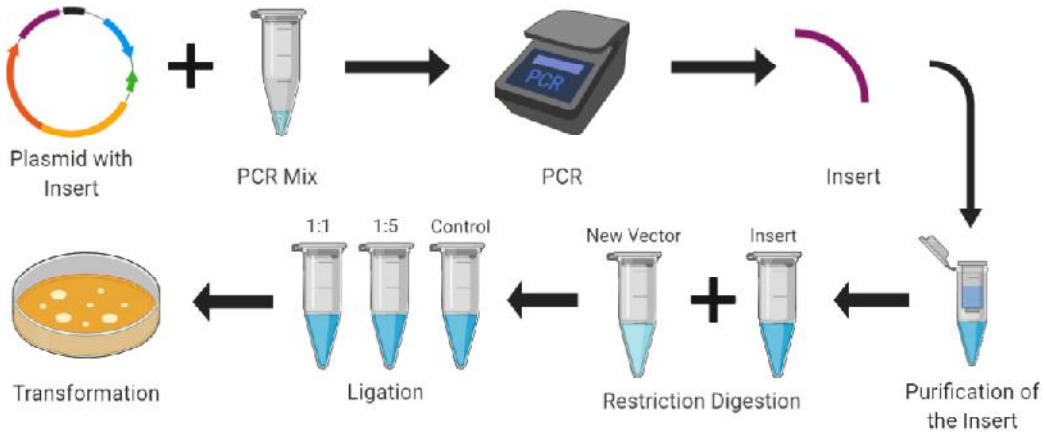


Figure 8 - Schematic process for cloning. First, the insert is amplified by PCR and purified, then both the insert and the vector are digested with restriction enzymes, then the insert is ligated with the vector and finally transformed in competent cells by electroporation. Creating using https://app.biorender.com/signin [17].

3.3.2.1. PCR

Initially, the PCR technique is applied to the initial vector that has the insert of interest. A 50 µL reaction mixture is prepared (Table 5).

Table 5 - Example of a PCR reaction. All buffers and enzymes used were purchased from New England BioLabs® [29].

Component	Reaction
5X Q5 Reaction Buffer	10 µL
10 mM dNTPs	1 µL
10 µM Forward Primer	2.5 µL
10 µM Reverse Primer	2.5 µL
Template DNA (1ng)	Variable
Q5 High-Fidelity DNA Polymerase	0.5 µL
Nuclease-Free Water	To 50 µL

After preparation of the 50 µL reaction it is necessary to place it in a thermocycler (ProFlex™ Thermo Fisher Scientific) and define the PCR parameters. It is necessary to define a temperature for denaturation of the DNA, then a set of 25 to 35 cycles is defined where DNA denaturation, primer binding and synthesis of new DNA repeatedly occur. Finally, there is a last extension and the program ends (Table 6).

Table 6 - Example of a PCR program.

Step	Temperature	Time
Initial Denaturation	98°C	30 seconds
25-35 Cycles	98°C	5-10 seconds
	50-72°C	10-30 seconds
	72°C	20-30 seconds
Final Extension	72°C	2 minutes
Hold	4-10°C	∞ time

3.3.2.2. Purification of the Insert

At the end of the PCR cycles, it is necessary to purify the insert. For this purpose, a NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel™) was used, and an agarose gel is used simultaneously to confirm the PCR results.

The DNA is then quantitated using a Nanodrop (NanoDrop™ DS-11 FC+, DeNovix) spectrometer based on the absorbance at 260 nm (1 absorbance unit = 50 ng/µL of double-stranded DNA).

3.3.2.3. Restriction Digestion

Once the insert is purified and the amount is suitable for further experiments, it is necessary to continue with the digestion of both the insert and the plasmid with restriction enzymes. A 50 µL mixture is prepared for the digestion of the insert and three reactions of 50 µL each for the digestion of the vector that will receive the insert (the vector is done in triplicate to compensate for the purification losses). In this mix it is necessary put 1µg of DNA, 10x Buffer diluted to 1X, this mean 1 µL of buffer per 10 µL of reaction, 1µL of each restriction enzymes and nuclease-free water to 50 µL. All the enzymes and buffers are purchased from New England BioLabs®.

Next, it is necessary to incubate the mix at 37°C for two hours and only for the vector add 1 µL (1 unit) of alkaline phosphatase (to prevent vector self-ligation) and incubate again for two hours.

Then it is necessary to purify the vector, for that an agarose gel was run and the vector was cut from the gel and purified using the NucleeSpin Gel and PCR Clean-up kit again.

3.3.2.4. Ligation

A mixture is prepared containing the insert and vector at different molar ratios 1 (Table 7). In this step, the enzymes and buffers are from Thermo Scientific®.

The mixtures are incubated overnight at 16°C and finally transformed in Clooni cells using the electroporation method (MicroPulser Electroporator BIO-RAD®).

Table 7 - Example of the Mix for insert ligation [30].

Ligation Mix	
Linear Vector	100 ng
Insert DNA	1:1 1:5 Control (only the vector)
10x T4 DNA Ligase Buffer	2 µL
T4 DNA Ligase	1 U
Water, nuclease-free	To 20 µL

3.3.3. EL222 protein

In order to carry out a photosensitivity study of EL222 protein, it was necessary to proceed with its production and purification. The whole protocol takes seven consecutive days with points where we can stop the process (Figure 9). For the EL222 protein production, the plasmid pHis-parallel1 (AmpR, Nt-6cHis, TEV site, see Annex), the host *E. coli BL21(DE3)* was used in LB medium with ampicillin or carbenicillin (a more stable analogue of ampicillin). The plasmid (#02, see Annex) encodes a fusion protein that contains a hexa-histidine tag and TEV cleavage site at the N-terminus and EL222 at the C-terminus.

Three buffer solutions are required for the execution of EL222 protein purification protocol, these being:

- Buffer A: Tris 50 mM, pH 8, NaCl 100 mM
- Buffer B: Tris 50 mM, pH 8, EDTA 0.5 mM, DTT 1 mM
- Buffer C: MES 50 mM, pH 6.8, NaCl 100 mM

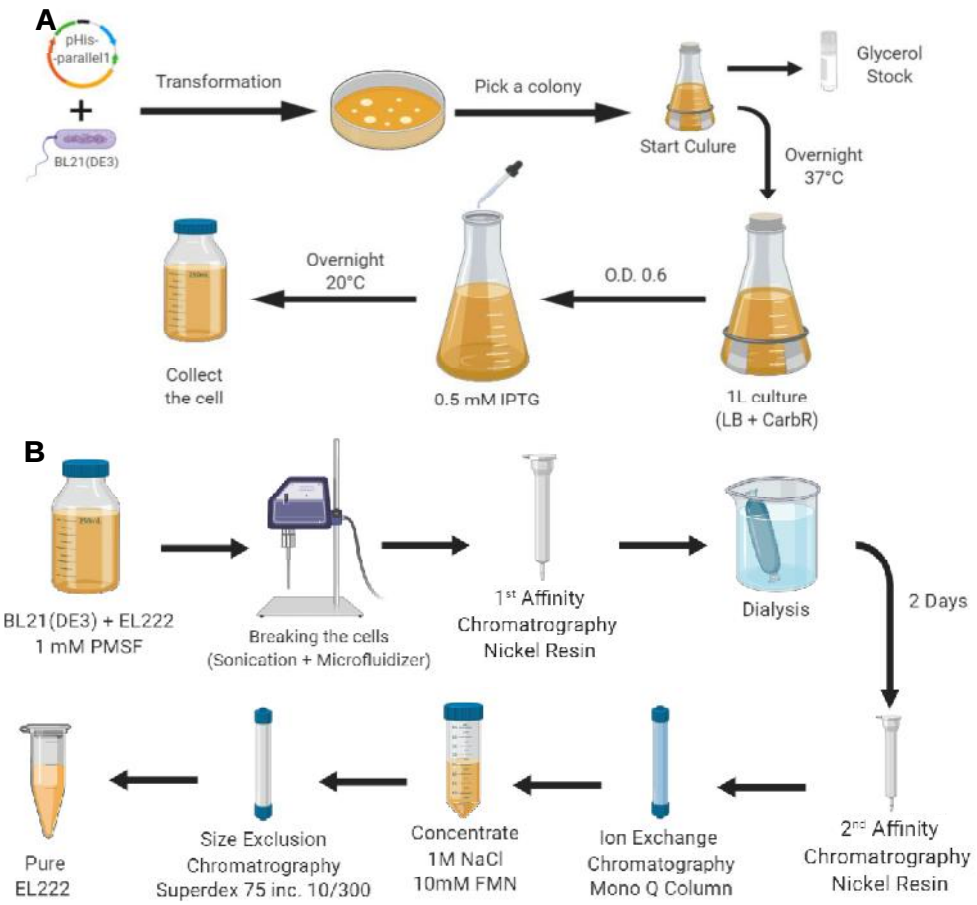


Figure 9 - Scheme of the process for the expression and purification of EL222 wild type protein. A) Transformation using competent cells, performing start culture and induction. B) Breaking of cells and all necessary steps for correct protein purification. Creating using https://app.biorender.com/signin [17].

3.3.3.1. Expressing the Protein (perform all steps with minimal ambient light)

1) Prepare the starter culture by inoculating a single colony from the plate (or glycerol stock) in 20 mL Luria Bertani (LB) medium containing Carbenicillin (Carb) antibiotics and incubate overnight at 37°C in the shaker (Thermo Scientific™ MaxQ™ 4000 Benchtop Orbital shakers).
2) Inoculate 10 mL of the starter culture in 1L of fresh LB medium containing Carb antibiotics and incubate in the shaker (Thermo Scientific™ MaxQ™ 5000 Floor-Model Shakers) at 37°C until the O.D.= 0.6.
3) Add 0.5 mM IPTG to induce protein expression.
4) Decrease the temperature to 20°C and incubated overnight at 250 rpm.
5) Next day, centrifuge the cells at 6000g for 15 minutes using the centrifuge (Avanti J-30I, Beckman Coulter Life Sciences).
6) Resuspend the cells in resuspension buffer and stored the cells at -80°C (only if is necessary to stop the process).

3.3.3.2. Breaking the cells

1) If the cell paste was less than 150 mL the sonication alone was used to lyse the cells. If the cell paste was more than 150 mL, the microfluidizer was used too.
2) Added protease inhibitor cocktail tablet or 1 mM PMSF.

3.3.3.3. Sonication

1) Use the sonicator (Qsonica Sonicator Q700, Fisher Scientific).
2) Donn't use the micro tip of the sonicator since it is not efficient for large volumens (more than 50 mL). Use only the screw.
3) The parameters were:
   a) Amplitude – 50;
   b) Process time – 1 min;
   c) Phase on time – 5 secs;
   d) Phase off time – 5 secs;
   e) Power should be 70W.

3.3.3.4. Microfluidizer

1) Use microfluidizer (EmulsiFlex-C3, AVESTIN).
2) Clean the microfluidizer and equilibrated with buffer A.
3) Pour the cell suspension.
4) Regulate the air pressure to 1500 bar.
5) If the suspension becomes viscous, sonicate and then microfluidize two times.
6) Centrifuge at 50,000g for 40 minutes at 4°C. (take a sample for SDS-PAGE)

3.3.3.5. 1st Affinity chromatography

1) Prepare three columns, each with 1mL beads (HisPur™ Ni-NTA Resin, Thermo Scientific).
2) Equilibrate the columns with buffer A.
3) Transfer the beads from column to a 50 mL falcon tube and centrifuge at 5000g for 4 minutes (Allegra X-30R Centrifuge, Beckman Coulter™).
4) Discard the supernatant buffer.
5) Incubat the Ni beads with protein on rocker (Rocker 35 EZ Large Capacity Lab Rocker, Labnet International) for 45 min at 4°C.
6) Add beads with the protein to the column. (take a sample for SDS-PAGE)
7) Wash the column with 60 mL resuspension buffer.
8) Wash the column with 20 mL washing buffer with 20 mM Imidazole. (take a sample for SDS-PAGE)
9) Elute the protein with 8 ml elution buffer with 500 mM Imidazole. (take a sample for SDS-PAGE)

3.3.3.6. Cleavage/Dialysis

1) Calculate the protein concentration by UV/Vis spectroscopy.
2) Add TEV enzyme. (take a sample for SDS-PAGE, before TEV)
3) Use 8 kDa pore size dialysis bag (Spectrum™ Spectra/Por™ 7 Membrane Tubing, Specrtrum™ 132131, fisher scientific).
4) Make 3 refolding buffers, all with 1 L and pH=8:
   a) 50mM Tris, 1 mM DTT and 0.5 mM EDTA. (overnight)
   b) 50mM Tris (morning)
   c) 20mM Tris (evening).
5) Perform the dialysis at 4 °C.

3.3.3.7. 2nd Affinity chromatography

1) Equilibrate the Ni beads (HisPur™ Ni-NTA Resin, Thermo Scientific) with buffer 20 mM Tris.
2) Put the protein.
3) Incubate the beads with protein on rocker for 30 min at 4°C in dark.
4) Collect the flow through using a centrifugal concentrator (ViVaspin® 15R Centrifugal Concentrators, sartorius). (take a sample for SDS-PAGE)

3.3.3.8. Final Purification of protein (NGC Discover™ 100 Chromatography System #7880010, BIO-RAD)

3.3.3.8.1. Ion Exchange chromatography (optional)

1) Use Mono Q column.
   a) Bed volume=1ml;
   b) Max column capacity=50mg protein;
   c) Flow rate=0.5ml/min.
2) Concentrate the protein volume until 8 to 10 mL.
3) Equilibrate the column with buffer 20 mM Tris, pH=8.

3.3.3.8.2. Centrifugal concentration

1) Add protein in the concentrator (ViVaspin® 15R Centrifugal Concentrators, sartorius).
2) Centrifuge (Allegra X-30R Centrifuge, Beckman Coulter™) at x6000 g until 5 ml.

3.3.3.8.3. Size exclusion chromatography (Superdex® 75 increase, GE Healthcare)

1) Add 10 mM FMN and 1000 mM NaCl to the protein.
2) Concentrate the protein volume until 1 mL.
3) Use MES buffer for the run.
   a. Total volume=24 mL;

b. Flow rate=0.75 mL/min;
c. Fraction size=1 mL (EL222 elutes typically in fraction 11)

3.3.4. Mutants of EL222

In order to make the expression and purification of the EL222 protein reproducible, a detailed protocol was elaborated that can be applied to both the wild type and the mutant versions of EL222.In this case, the expression plasmid (#106, see Annex) encodes a fusion protein with EL222 at the N-terminus and an intein-CBD-12Histidine tag at the C-terminus.

For the incorporation of the non-canonical amino acid 4-cyanophenylalanine (CNF) at pre-defined positions, the aforementioned protocol suffers the following changes, illustrated also in Figure 10:

- Use Terrific Broth medium.
- The transformation of *E. coli BL21(DE3)* cells is performed with two plasmids, one containing the mutated EL222 gene and resistance to ampicillin (e.g. plasmid #106) and the other containing an orthogonal suppressor aminoacyl tRNA synthetase/tRNA pair specific for CNF and resistance to spectinomycin (plasmid #72, see Annex);
- Non-canonical amino acids (CNF) are added to the culture medium at a concentration of 1 mM when O.D. = 0.4;
- Induction of the culture with 1 mM IPTG is performed after 30 min of the addition of the non-canonical amino acid;
- In the first affinity chromatography with nickel (Ni) a buffer of 40 mM imidazole is used for washing.
- The cleavage is performed by the intein tag andtakes two days. The protein was mixed with DTT (Buffer for cleavage: 50 mM tris-HCl pH=9, 50mM DTT).
- The dialysis is made after the cleavage.

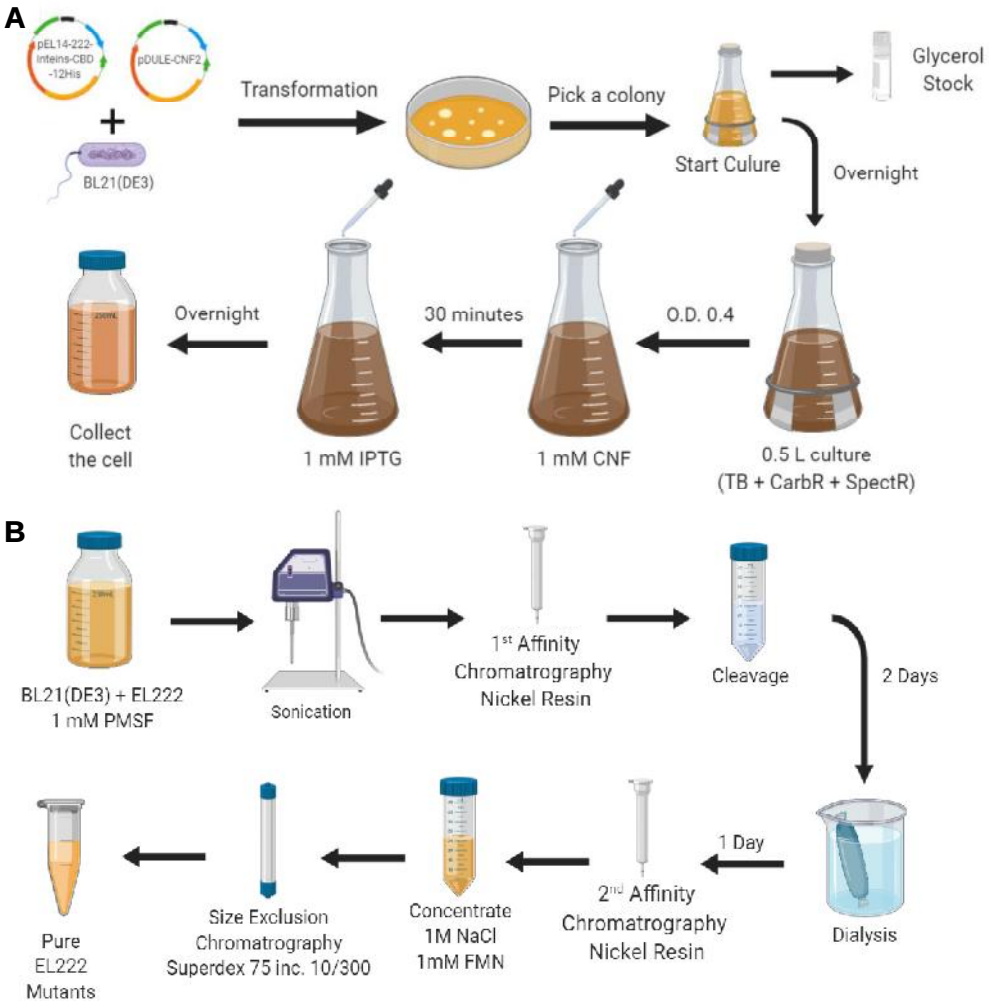


Figure 10 - Scheme of the process for the expression and purification of EL222 mutant proteins carrying the non-canonical amino acid CNF. A) Transformation using competent cells, performing start culture, added non-canonical amino acid and induction. B) Breaking of cells and all necessary steps for correct purification of mutant proteins. Creating using https://app.biorender.com/signin [17].

3.3.5. CarH full-length protein

The CarH full-length protein purification process is a protocol that takes place in just one day due to protein instability. Figure 11 shows schematically the steps required to obtain this protein, however it should be noted that light does not interfere with the protein until the cofactor is added.

For the purification of CarH full-length protein the below buffers were necessary:

- Buffer A: 300 mM NaCl, 50 mM Sodium Phosphate, pH=7.
- Buffer B: 300 mM NaCl, 50 mM Sodium Phosphate, 10 mM imidazole, pH=7.
- Buffer C: 300 mM NaCl, 50 mM Sodium Phosphate, 150 mM imidazole, pH=7.

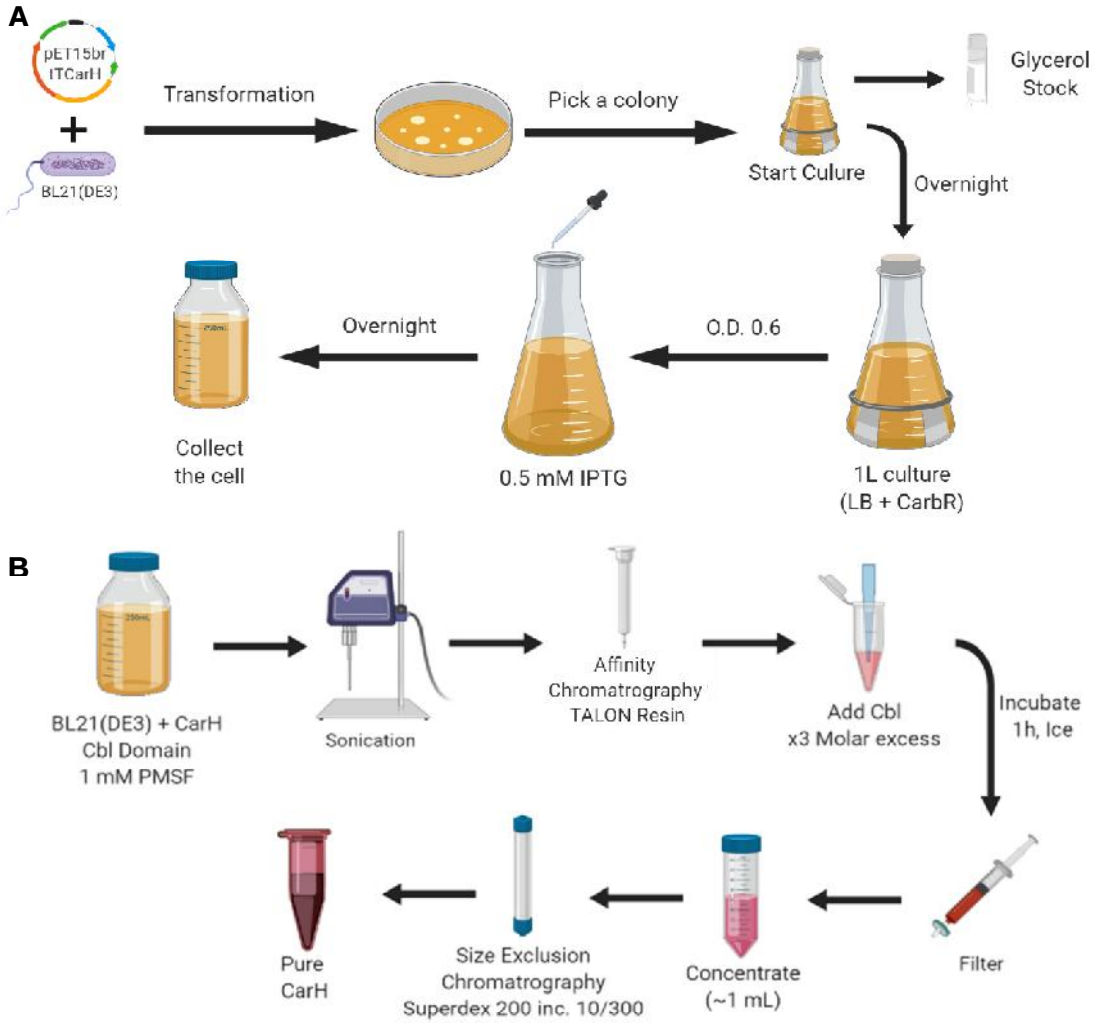


Figure 11 - Scheme of CarH full-length protein expression and purification process. A) Transformation using competent cells and performing start culture. B) Breaking of cells, add the cofactor and the necessary steps for purification. Creating using https://app.biorender.com/signin [17].

3.3.5.1. Expressing the Protein

1) Use the plasmid pET15brtTCarH (#059, see Annex) and *E. coli BL21 (DE3)* competent cells.
2) Inoculate 20 mL Luria Bertani (LB) medium containing carbenicillin antibiotics (Carb) with one colony and incubate overnight at 37°C in the shaker (Thermo Scientific™ MaxQ™ 4000 Benchtop Orbital shakers).
3) Inoculate 10 mL of the starter culture in 1L of fresh LB medium containing Carb antibiotics and incubate in the shaker (Thermo Scientific™ MaxQ™ 5000 Floor-Model Shakers) at 37°C until the O.D.= 0.6.
4) Added 0.5 mM IPTG to induce the culture.
5) Set the temperature to 25°C and incubate overnight at 250 rpm.
6) Next day, centrifuged the cells at 6000g for 15 minutes using the centrifuge (Avanti J-30I, Beckman Coulter Life Sciences).

7) Resuspend the cells in buffer A and store the cells at -80°C (only if is necessary to stop the process).

3.3.5.2. Breaking the Cells

1) Added protease inhibitor cocktail tablet or 1 mM PMSF.
2) Use the sonicator (Qsonica Sonicator Q700, Fisher Scientific).
3) Don't use the micro tip of the sonicator. Used only the screw.
4) Use the parameters below:
   a) Amplitude-50;
   b) Process time-1min;
   c) Phase on times-5secs;
   d) Phase off time-5secs;
   e) Power should be ~70W.

3.3.5.3. Affinity Chromatography

1) Prepare one column, each with 1mL Talon resin (TALON® Metal Affinity Resin, TaKaRa).
2) Equilibrate the columns with buffer A.
3) Transfer the Talon resin from column to a 50 mL falcon tube and centrifuge at 5000g for 4 minutes (Allegra X-30R Centrifuge, Beckman Coulter™).
4) Discard the supernatant buffer.
5) Incubate the Talon resin with protein on rocker (Rocker 35 EZ Large Capacity Lab Rocker, Labnet International) for 20 min at 4°C in dark.
6) Add Talon resin with the protein to the column. (take a sample for SDS-PAGE)
7) Wash the column with 50 mL of buffer A.
8) Wash the column with 20 mL of buffer B. (take a sample for SDS-PAGE)
9) Elute the protein with 5 mL of buffer C. (take a sample for SDS-PAGE)

3.3.5.4. Final Purification

1) Quantify the protein using the Nanodrop (NanoDrop™ DS-11 FC+, DeNovix).
2) Added three-fold (3X) molar excess of AdoCbl over protein.
3) Incubate for 1h in ice.
4) Filter with a syringe.
5) Concentrate to less than 1 mL.
6) Centrifuge the sample at maximum speed, 16°C.

3.3.5.4.1. Size Exclusion Chromatography (NGC Discover™ 100 Chromatography System #7880010, BIO-RAD)

1) Use Superdex 200 increase 10/300 GL, GE Healthcare with Buffer A.

3.3.6. CarH Cobalamin Domain protein

In the case of CarH containing only the cobalamin-binding domain (residues 76 to 285, plasmid #59, see Annex), the protocol is very similar to that of full-length protein (Figure 12). Only three steps are different:

- During affinity chromatography the resin used is Chitin instead of Talon.
- Protein cleavage occurs during the affinity chromatography.
- Added a step for dialysis against buffer A and make the step in one day.

For this protein, the purification protocol is performed on two or three consecutive days..

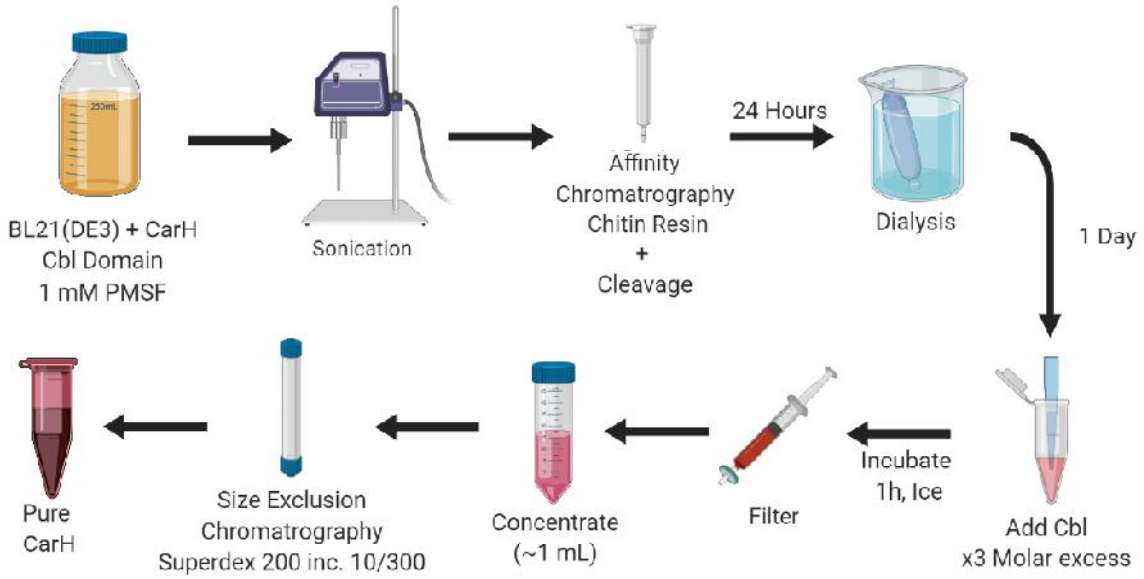


Figure 12 - Representative scheme of the CarH cobalamin binding domain purification process. Creating using https://app.biorender.com/signin [17].

3.3.6.1. Affinity Chromatography/ Cleavage

Buffers for the chitin resin (Chitin Resin, New England BioLabs®) column:

- Column buffer (Buffer 1): 300 mM NaCl, 50 mM sodium phosphate (pH 7).
- Cleavage Buffer (Buffer 2): 20 mM Na-HEPES, 300 mM NaCl, 50 mM DTT, 1mM EDTA; 0,05% Tween-20, pH 8,5
- Regeneration buffer (Buffer 3): 300 mM NaOH

1) Column Preparation
   a) Put 10 mL of “wet” resin (5mL of “dry” resin).
   b) Washed with water (50mL).
   c) Washed with Column Buffer (20mL).
2) Loading
   a) Put the protein (supernatant) and resin in a falcon tube.
   b) Add 0,05% Tween-20.
   c) Incubate 2h at 4°C.
   d) Drained the Flow Through (keep).

3) Washing
    a) Wash with Column Buffer.
4) Adding Thiol
    a) Added 5 mL of Cleavage Buffer.
    b) Collect the Flow Through (very fast, keep just in case).
    c) Add 5 mL of Cleavage Buffer.
    d) Mix in the column (resuspend).
5) On-column Cleavage
    a) Incubated 24h at 23°C.
6) Elution
    a) Collect the flow-through (5mL of protein).
    b) Added 2 mL of Column Buffer.
    c) Collect again in the same falcon (final protein volume ~ 7 mL).

3.3.7. Biophysical characterization of proteins

At the Institute of Biotechnology, the researchers have at their disposal a wide range of equipment for their experiments. Part of the equipment is available at the Center of Molecular Structure (CMS) facility on a "service" basis. It means that the user (in this case, myself) sends the sample and gets the results without direct participation in the experiments themselves (which are performed by highly qualified CMS personnel)

All results presented in chapter 4 without associated protocol were performed by colleagues from CMS.

Following are the protocols of the experiments performed by the intern.

3.3.7.1. UV-Vis spectroscopy

The first experiment to be performed on both proteins under study was to monitor the changes in the chromophore upon light irradiation by UV-Vis spectroscopy. To perform the experiment, it is first necessary to measure the blank using the protein buffer, then prepare a 70-fold dilution in a cuvette (UV-Cuvette micro, centre height 8.5 mm, dimensions 12.5 x 12.5 x 45 mm, BRAND®) using the respective buffer of each protein. In this first measurement, the protein has not yet been irradiated, thus representing the dark state. The absorbance corresponding to protein (Abs 280) and cofactor (FMN - Abs 450, AdoCBL - Abs360) is used to calculate the protein concentration (Table 8). Such spectrum represents also the dark state. Then the protein is irradiated with the appropriate light and then a second reading, corresponding to the tlight state, is made.

Table 8 - Extinction Coefficient and wavelengths of proteins.

	Extinction Coefficient	**Wavelengths (nm)**
CarH Full-Length	37930	360
CarH cbl domain only	20970	360
EL222	13000	450

For protein irradiation, the following center wavelengths emitted from LEDs (THORLABS) are used:

- 617 nm (red light);
- 530 nm (green light);
- 455 nm (blue light);
- 365 nm (ultraviolet light).

During the experiment, the LEDs were adjusted to apply the same 34 mJ energy (power x time) so that a reliable comparison becomes possible. Nanodrop is used for all measurements (NanoDrop™ DS-11 FC+, DeNovix).

3.3.7.2. Microscale Thermophoresis – MST

The MST experiment was performed only for CarH protein. In this experiment, the DNA concentration remains constant and the protein concentration varies.

This experiment requires a small amount of protein (60 μM) and a suitable buffer (100 mM KCl, 0,00025 mM Tris pH = 8, 1 mM DTT, 10 mM $MgCl_2$ and 0.1% Pluronic or 0.05% Tween-20, pH = 8).
First, sixteen 1:1 dilution series of the protein are prepared in the appropriate buffer (starting at a concentration of 30 μM) and 10 nM Cy5-DNA is added to each dilution. After preparation of the sixteen samples, they are placed in capillaries, which are in turn placed in the MST equipment (Monolith NT.115, Nanotemper) for measuring the dark state.The most suitable parameters for CarH protein are:

- Fluorescence excitation power - 20%;
- MST power - 20%;
- Laser time - 5 seconds.

After the dark state is measured the capillaries containing the sample are irradiated with the 530 nm LED for 20 seconds (to obtain the light state) and a new measurement is taken. With this experiment the affinity between protein and DNA is expressed as K_D or dissociation constant (the lower the K_D the higher the affinity. K_D values are obtained in both dark and light states.

3.3.7.3. Fourier-transform Infrared Spectroscopy – FTIR

Taking reliable Fourier-transform infrared (FTIR) spectra depends on some critical steps:

- Correct preparation of equipment (VERTEX 70V, BRUKER);
- Changing the protein buffer to D_2O to eliminate water interference in the final spectrum (the H_2O absorption band overlaps with the protein Amide I band);
- Keep temperature constant (20 °C) so that the spectra are not affected by fluctuations in surrounding environment.

It is necessary to pay attention to whether the protein measurement is being done in the dark or light state and to keep the room in the right conditions (such as turning off or turning on the room lights).

Before starting each measurement, make sure that the equipment is correctly prepared and that the transmission cell to be used is correctly placed. A thermostated calcium fluoride (CaF_2) cell of 25 µm path-length was used (Specac). Then start a series of measurement, each measurement has 100 to 1000 scans, depending on the protein concentration (low protein concentrations require more scans to reach a given level of signal-to-noise ratio). For each measurement, two spectra must be taken: the reference spectrum and the sample spectrum. The reference spectra can be buffer or a protein sample. 30 µl are needed. For the dark state (three repetitions), the protein is kept protected from light. For the light state, the sample is irradiated with a LED emitting green light (λ = 530 nm) for CarH protein or blue light (λ = 455 nm) for EL222 protein, and measured again (three times).

Briefly, the four spectra for a complete protein analysis are:

- Dark minus Buffer - Background: Buffer, protein measurement in dark state (absolute spectra);
- Dark minus Dark - Background: Dark state protein, control for instrument fluctuations in the dark (difference spectra);
- Light minus Dark - Background: Dark state protein, protein measurement in light state (difference spectra);
- Light minus Light - Background: light state protein, control for instrument fluctuations in the light (difference spectra).

4. RESULTS AND DISCUSSION

4.1. Preparation of proteins

4.1.1. LOV and HTH domain of EL222

In order to better understand the EL222 protein, two clones were produced, one with only the DNA-binding domain (HTH domain) portion and another with the LOV domain only. Table 9 represents the final name given to the new plasmids created at the end of the cloning process, as well as the vector used, the insert, the restriction enzymes and, finally, the antibiotic resistance conferred by the vector.

Table 9 - Cloning of LOV and HTH domains.

Final Plasmid	Vector	Insert	Restriction enzymes	Antibiotic resistance
pET30-His12-Gb1-EL222(LOV)-CfaN	#89	LOV	NdeI + EcoRI	Kanamycin
pET30-CfaC-EL222(HTH)-Intein-CBD_G2A	#89	HTH	NdeI + EcoRI	Kanamycin

4.1.2. Mutagenesis

After proceeding with the entire site-directed mutagenesis protocol, the transformation was done by electroporation using the DNA mixture and *E. coli* clooni cells. For transformation, 50 µl cells and 1 µl of mutagenesis reaction were used. All plasmids for the EL222 protein mutants carry an ampicillin (or carbenicillin) antibiotic resistance cassette and plates with that same antibiotic were used.

Plates were incubated at 37 °C overnight and the next morning colonies were observed. Many colonies were obtained, as can be seen from Figure 13.

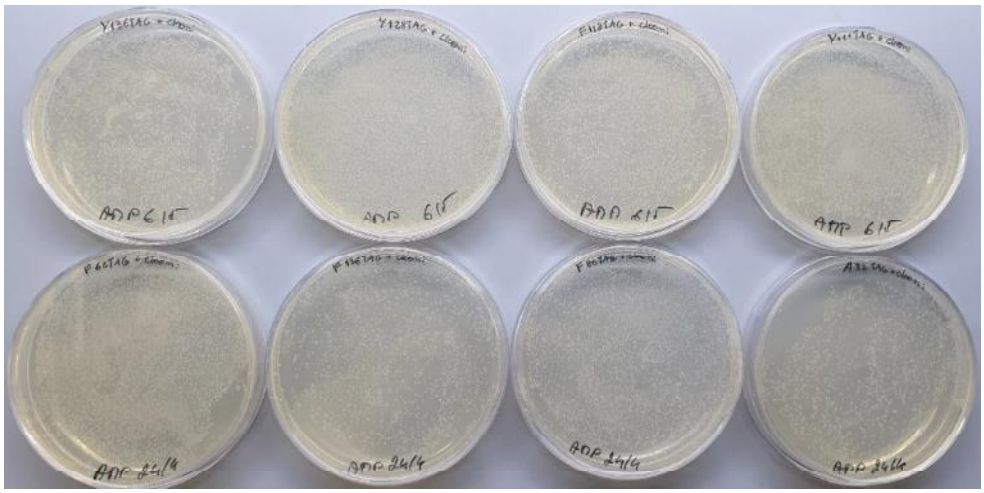


Figure 13 - Plates used for electroporation transformation of EL222 protein mutants. 50 µl of electrocompetent E. coli Clooni cells and 1 µl of the mutagenesis mixture were used. The plates remained at 37ºC overnight.

Then four colonies were randomly taken from each plate and the indicated plasmid purification kit was applied. Once the plasmid was obtained, it was quantified and sent for sequencing (Eurofin Genomics). After a few days of waiting, the sequencing results were available, enabling the identification of which plasmids contained the desired mutation. In Table 10 it is possible to visualize the created mutants and whether they had the expected sequence or not.

Table 10 - List of generated EL222 mutants.

Mutants	Mutation found by sequencing
F62TAG	✓
F80TAG	✓
F118TAG	✓
F137TAG	✓
Y68TAG	✓
Y111TAG	✓
Y128TAG	✓
Y136TAG	✓

It was found that the protocol for mutagenesis was succesfully applied and correct plasmids were obtained for the eight desired mutants. Given this, one could begin to express and purify the corresponding proteins.

4.2. Protein expression and purification

4.2.1. EL222 protein

4.2.1.1. EL222 Wild Type

During purification of wild type EL222 protein, several samples were taken at strategic points of the process to determine if the purification was being performed well. At the end, all samples are run on an SDS-PAGE gel and the evolution of the purification process was observed (Figure 14).

The first sample collected was the supernatant (S), after the cells are broken down and the mixture is centrifuged. The supernatant is then placed on the column to perform the first affinity chromatography and the flow-through (F) was collected. During the first affinity chromatography, wash (W) and elution (E) samples were also collected. Finally, a sample was taken after tag cleavage (AC), for this 0.36 mL of TEV was added.

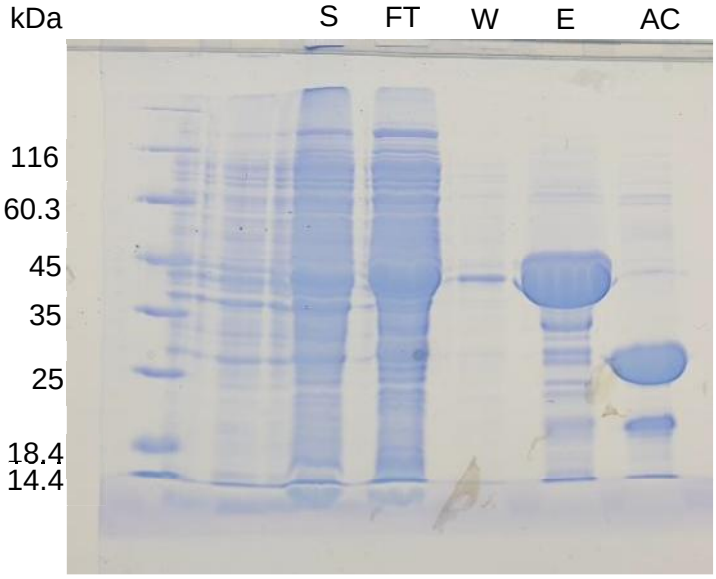


Figure 14 - SDS-PAGE image of the purification process of the EL222 wild type protein. S – supernatant; FT – Flow through; W – wash; E – Elution; AC – After Cleavage.

Looking at the SDS-PAGE image shown in Figure 14, it can be said that the purification process of wild type EL222 protein went as predicted until the tail-cut step of the purification (AC). In the supernatant and flow-through samples, many contaminating proteins are present, as would be expected. In the washing step, it can be seen that there is a small loss of the protein concerned, but nothing too significant. In the elution we find that we have the protein present in an acceptable amount, but still has many contaminants. Finally, after cleavage, it can be seen that it has been properly performed as the protein changes from 34 kDa (wild type EL222 protein plus purification tags) to 25 kDa (wild type EL222 protein without extra tags).

To obtain a purer protein, ion exchange chromatography was performed. In Figure 15 it is possible to observe the obtained chromatogram, and as expected the separation occurred quite well since there is little overlap of the peaks. The peak collected for use in the next step was the second (34 - 37 mL).

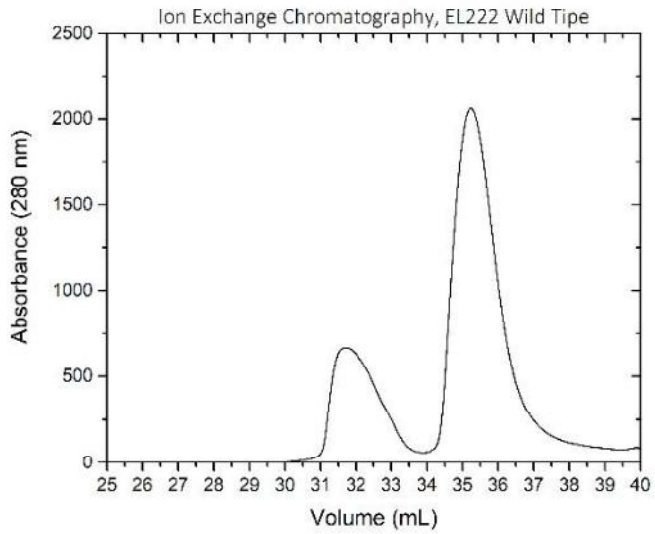


Figure 15 - Ion-exchange chromatogram of wild type EL222 protein.

Following ion-exchange chromatography, size-exclusion chromatography was performed to obtain the protein in the purest possible state. In Figure 16 it is possible to observe the obtained chromatogram, in which an isolated and well-defined peak was obtained, which suggests that a single species is present in the sample.

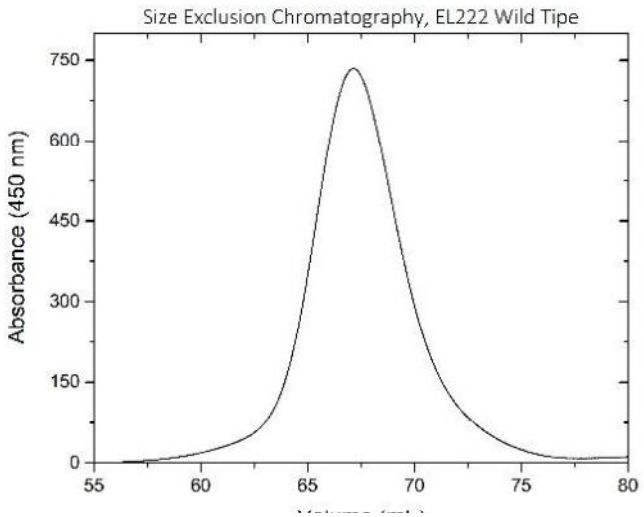


Figure 16 - Size-exclusion chromatogram of wild type EL222 protein.

Fractions 3 to 10 of size exclusion chromatography were collected, and samples were prepared for SDS-PAGE to verify protein purity. After running the gel (Figure 17), it was seen that. the fractions were quite pure, although some had a minimal amount of contaminants.

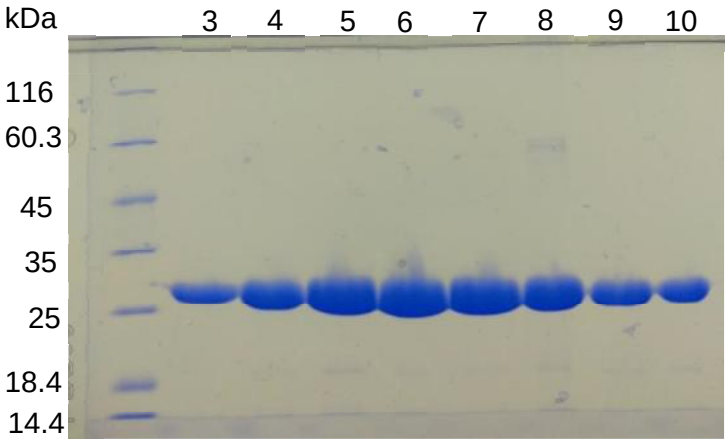


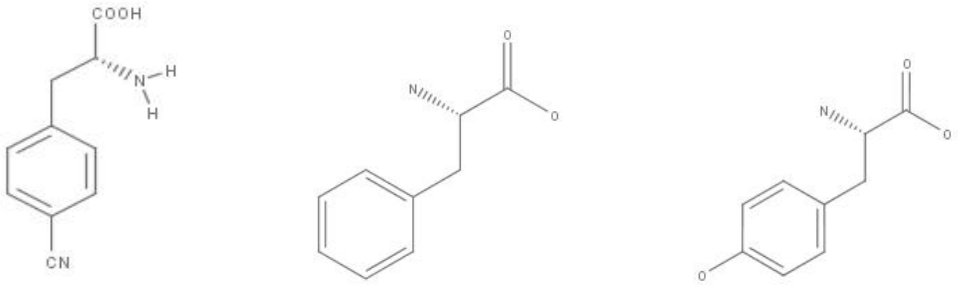
Figure 17 - SDS-PAGE imagel of the fractions collected after size exclusion chromatography of the EL222 wild type protein.

To finalize the protein purification process, and despite minimal contamination, fractions 3 to 10 were collected, mixed and brought to 600 µl and a final concentration of 4.89 mM. Finally, the concentrated protein was stored at -80 °C for future studies.

4.2.1.2. EL222 CNF Mutants

For further study of the EL222 protein eight mutants were created using site-directed mutagenesis. Each variant has a mutation at a predetermined position to determine which exact sites of this domain undergo a change in conformation, what structural change and at what time point after the protein is irradiated with blue light.

Labeling was done using the non-canonical amino acid 4-cyano-L-phenylalanine or CNF. This specific non-canonical amino acid was used because the vibrational frequency of its cyano group is sensitive to the surrounding environment (polarity, electrostatic, etc.) and has a structure very similar to the canonical amino acids tyrosine and phenylalanine (Figure 18). Due to these facts, it is expected that this mutation will not cause major changes in protein functioning and will be identifiable by the FTIR technique, thus determining if there is any change in the mutation zone when it is exposed to light.



CNF Phenylalanine Tyrosine

Figure 18 - Structure of CNF, phenylalanine and tyrosine.

The four native phenylalanine and four tyrosine residues of EL222 (all residing in the LOV domain) were mutated to CNF. Of the eight mutants created all were expressed using *E.coli* BL21 (DE3) cells in 0.5 L of TB medium, however, due to the short time of internship, it was not possible to purify all mutants. Table 11 provides information on all mutants that have been expressed and purified. Notice that F stands for phenylalanine and Y is short for tyrosine, and then the number indicates the position of the amino acid in the sequence. Therefore, F62CNF means that the native phenylalanine at position 62 has been replaced by CNF using amber suppression technology.

Table 11 - EL222 mutant proteins containing the non-canonical amino acid CNF.

Mutants	Expressed Mutants	Purified Mutants
F62CNF	✓	✓
F80CNF	✓	✓
F118CNF	✓	✓
F137CNF	✓	✓
Y68CNF	✓	✓
Y111CNF	✓	✓
Y128CNF	✓	×
Y136CNF	✓	×

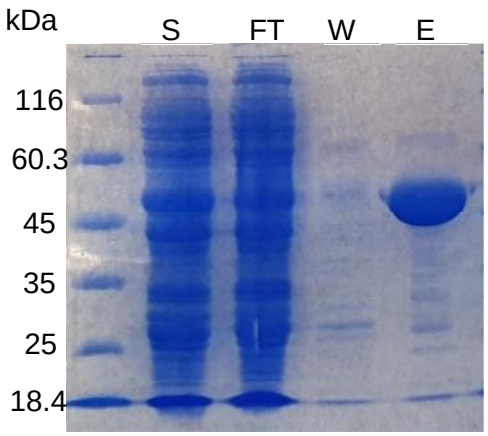
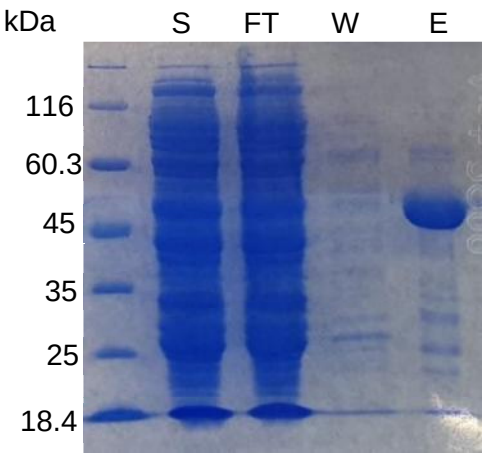
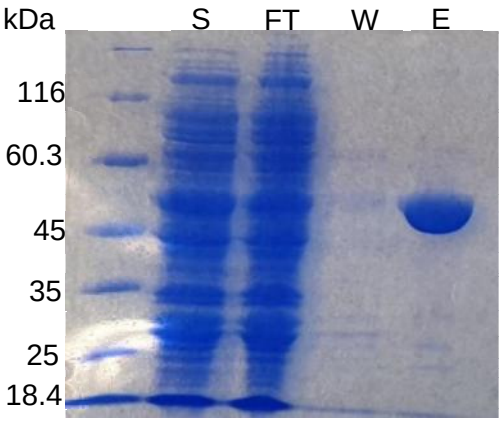
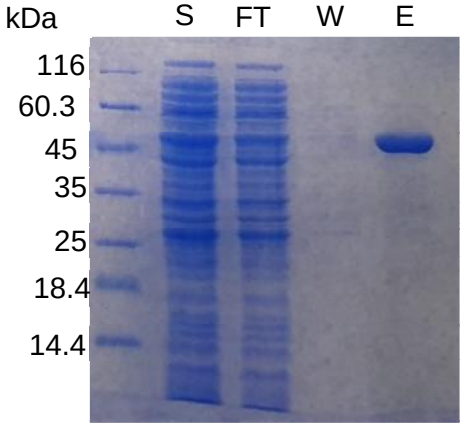
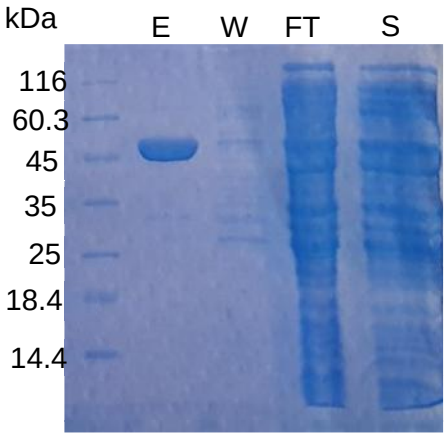
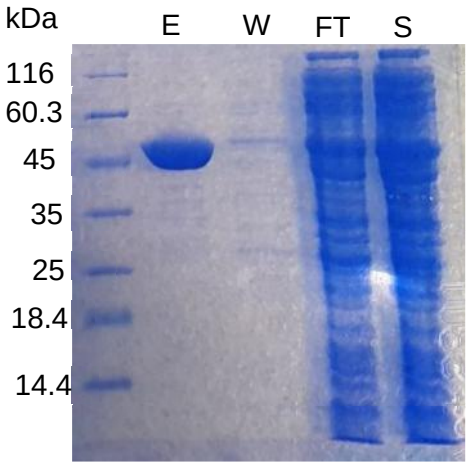
At the end of the purification of the six mutants, a final volume between 60 and 80 µL, and a concentration between 0.33 and 1.87 nM were obtained. Table 12 shows the exact volume obtained for each mutant and its final concentration calculated from the absorbance measured by either UV-Vis or infrared spectroscopies. Notice that UV-Vis reports on the FMN concentration while infrared is sensitive to the total protein concentration.

Table 12 - Obtained amount of EL222 mutant proteins.

Mutants	Volume (µL)	Final Concentration (mM)	
		UV-Vis	Infrared
F62CNF	80	1.46	1.34
F80CNF	60	0.27	0.33
F118CNF	50	0.88	1.12
F137CNF	70	1.12	0.65
Y68CNF	60	3.21	1.8
Y111CNF	60	3.01	1.87

As a control for the purification process, several samples were taken at various stages. The first sample taken was the supernatant (S), then flow through (FT), washing and finally elution of the affinity chromatography column. Table 13 shows all SDS-PAGE gels performed for the studied mutant protein. By observing the SDS-PAGE images, it is possible to state that the purification process went as expected and mutants of the appropriate size (52 kDa before cleavage) were obtained.

Table 13 - SDS-PAGE to monitor purification of EL222 mutant proteins.

F62CNF	F80CNF
	
F118CNF	**F137CNF**
	
Y68CNF	**Y111CNF**
	

All mutants were subjected to size-exclusion chromatography to obtain proteins in the purest possible form. Figure 19 shows the chromatograms obtained, where it is possible to observe that all mutants were eluted from the column in the same volume; however some had a lower absorbance value which indicate a lower concentration and therefore a lower expression yield.

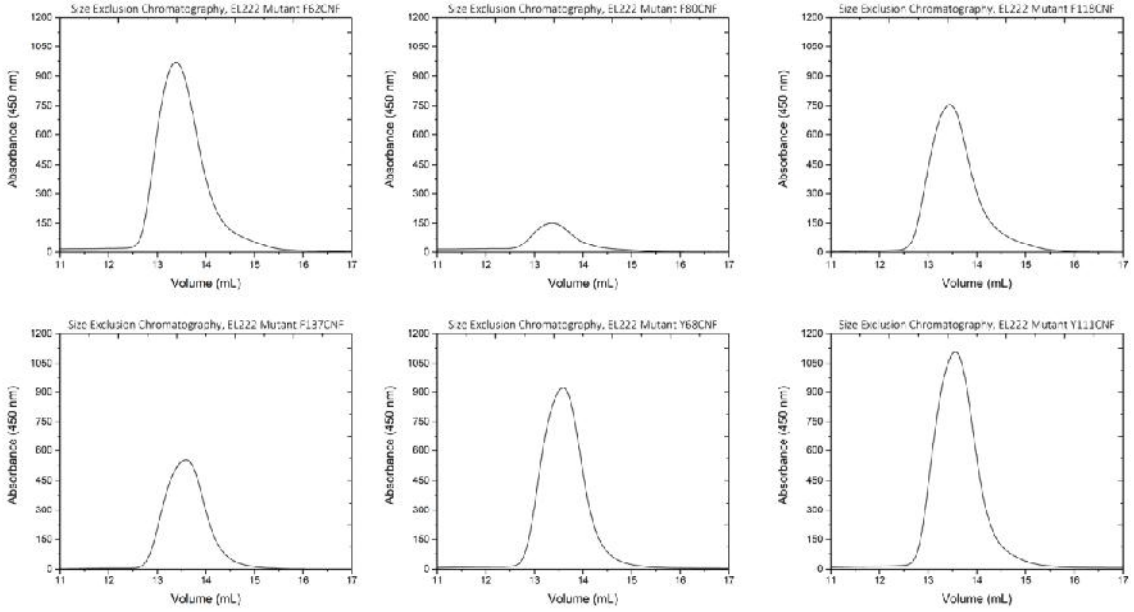


Figure 19 - Size-exclusion chromatograms of EL222 CNF mutants.

At the end of the purification of the mutated EL222 protein variants, it is possible to state that the F80CNF and F137CNF mutants were the ones with the lowest concentration, which may indicate that the mutation is not as harmless as expected and may cause some changes in the synthesis and/or folding of the protein *in vivo*.

4.2.2. CarH proteins

4.2.2.1. CarH full-length

Initially the CarH protein expression and purification process was started using *E. coli BL21(DE3)* in 3 liters of LB media. After optimization of the purification process and as soon as it was possible to produce and purify the protein correctly, the process scale was increased so that the amount of protein obtained was higher. This protein batch was obtained from 8 liters of LB media.

To better control the purification process, samples were taken at strategic points throughout the protocol to ultimately make an SDS-PAGE. Figure 20 shows the gel made where a sample taken from the supernatant, flow-through, washing and elution of the affinity chromatography column was placed and finally a sample of the final protein obtained (purified and concentrated).

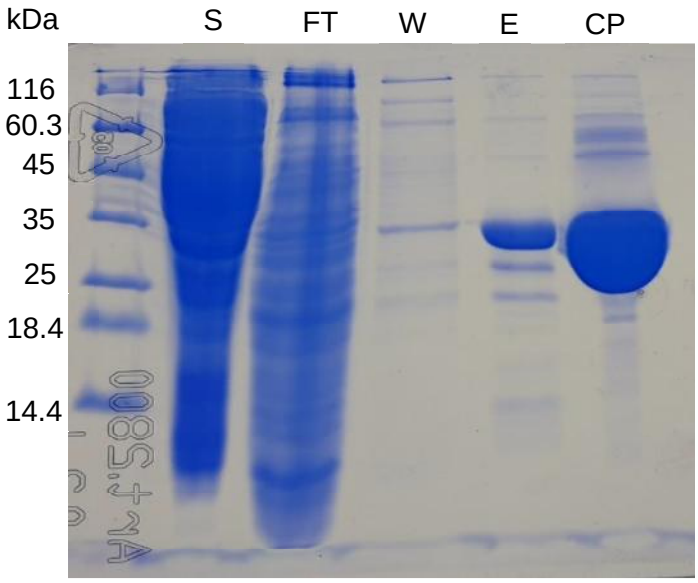


Figure 20 - SDS-PAGE image showing the purification process of CarH Full Length protein. S- Supernatant; FT – Flow through; W – Washing; E – Elution; CP – Final protein.

In figure 20 it is possible to observe that the expected protein was obtained because it presents the correct size of the protein CarH plus the His-tag for the purification (33 kDa), it is also possible to observe that the final protein presents some contaminating proteins but in quantities so low that they were not taken into account. In the case of CarH full-length protein, the purification tail was not cleaved because there is a potential thrombin cleavage site in the CarH sequence.

In the size-exclusion chromatography step, the chromatogram present in Figure 21 was obtained and, as expected, the separation was well performed because the peaks show a good separation and are not superimposed. The first elution peak (13 - 15 mL) corresponds to CarH protein in tetrameric form (non-irradiated and properly assembled) and the second peak (16 - 17 mL) in monomeric form (inadvertent light irradiation or improper reconstitution with AdoCbl). This is because it is impossible for the operator to ensure that the protein is not exposed to light (even if not directly) during the entire purification process.

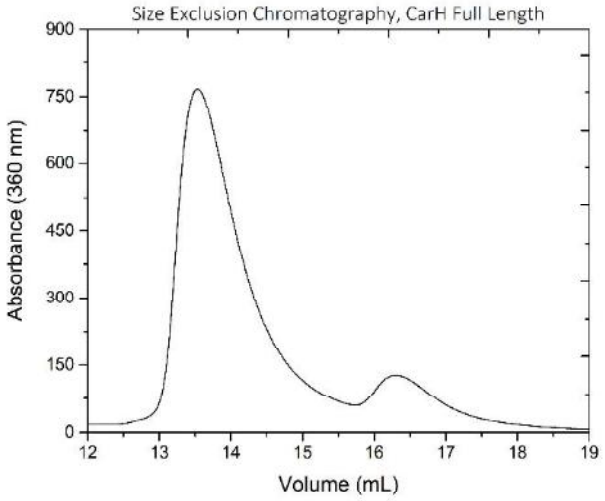


Figure 21 - Size exclusion chromatogram of CarH full-length protein.

The final volumes and concentrations obtained at the end of purification can be seen in Table 14.

Table 14 - Obtained amount of CarH full-length proteins.

Protein (state)	Volume (µL)	Concentration (mM)	
		UV-Vis	Infrared
Monomer	800	0.03	0.09
Tetramer	150	0.7	1.25

The final samples of the purified protein were not stored at -80 °C but instead they were subjected immediately to the planned studies.

4.2.2.2. CarH cobalamin binding domain

The CarH cobalamin domain protein is a truncated version of full-length CarH protein mutant that has only the cofactor binding domain, in this case AdoCbl. Hence, this protein is not expected to bind DNA since it does not have the domain for it.

CarH cobalamin domain protein was initially produced using *E. coli BL21 (DE3)* in a culture containing 1L LB medium. However, the process was optimized and the last batch was produced using 3 L of medium. This protein was more stable and showed higher expression yield than CarH full-length protein.

During the protein purification process, several samples were taken at key stages to monitor its progress. At the end of the purification procedure, the supernatant, the flow-through, wash, cleavage buffer wash, protein after cleavage, protein after dialysis, a column regeneration sample and finally the pure protein were analysed by SDS-PAGE. Figure 22 shows the SDS-PAGE image performed after protein purification.

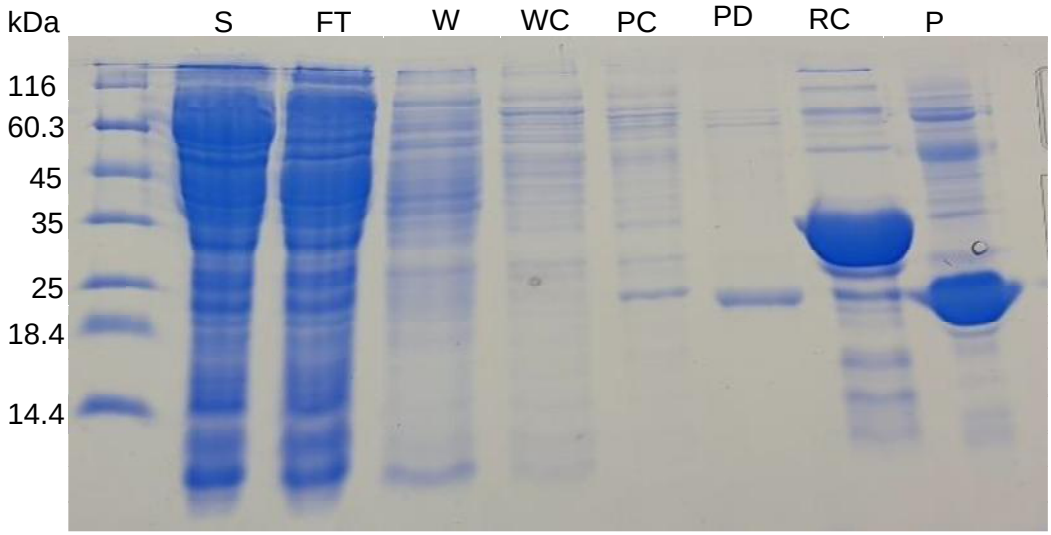


Figure 22 - SDS-PAGE image showing the purification process of CarH cobalamin domain protein. S – Supernatant; FT – flow through; W – Wash; WC – Cleavage buffer wash; PC – Protein after cleavage; PD – Protein after dialysis; RC – Column regeneration; P – Final protein.

Looking at the SDS-PAGE gel of Figure 22, it can be seen that an error in the protein cleavage occurred as much of the protein was retained in the column possibly because not having cleaved off its purification tag. This error could have been quite problematic because in the end the amount of protein might have not been enough to continue the tests. However, this was not the case. This error was due to incorrect preparation of cleavage buffer.

To obtain a protein as pure as possible, the last step of purification was size-exclusion chromatography. Figure 23 shows the chromatogram obtained for CarH cbl domain protein with the first elution peak (14 - 16 mL) showing the protein in the tetrameric state (non-irradiated and properly assembled) and the second peak (17 - 19 mL) showing the protein in monomeric state (inadvertent light irradiation or improper reconstitution with AdoCbl). The two peaks are well separated.

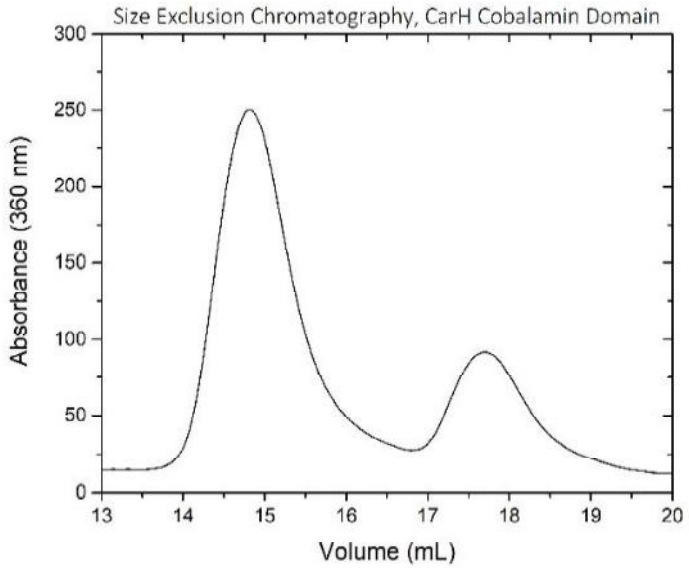


Figure 23 – Size-exclusion chromatograpm of CarH cbl domain protein.

At the end of purification 80 µl of the monomer protein with a concentration of 0.24 mM and 30 µl of the tetrameric protein with a concentration of 1.30 mM were obtained. These values can be seen in table 15.

Table 15 - Obtained amount of CarH cbl domain proteins.

Protein (state)	**Volume (µL)**	**Concentration (mM)**	
		Uv-Vis	Infrared
Monomer	80	0.17	0.24
Tetramer	30	0.6	1.30

The final samples of the purified protein were not stored at -80°C but instead they were subjected immediately to the planned studies.

4.3. Characterization of EL222 mutants

After expressing and purifying the EL222 protein mutants, their spectra were obtained using UV-Vis and FTIR spectroscopies.

4.3.1. UV-Vis spectroscopy.

The spectra obtained for the six EL222 CNF mutants were obtained using a quartz cuvette of 1 cm path-length (Nanodrop) and for all of them a 70-fold dilution was performed.

In Figure 24 it is possible to observe the dark state spectra by UV-Vis. As expected, all mutants have a very similar spectrum and the main differences arise from the different protein concentrations.

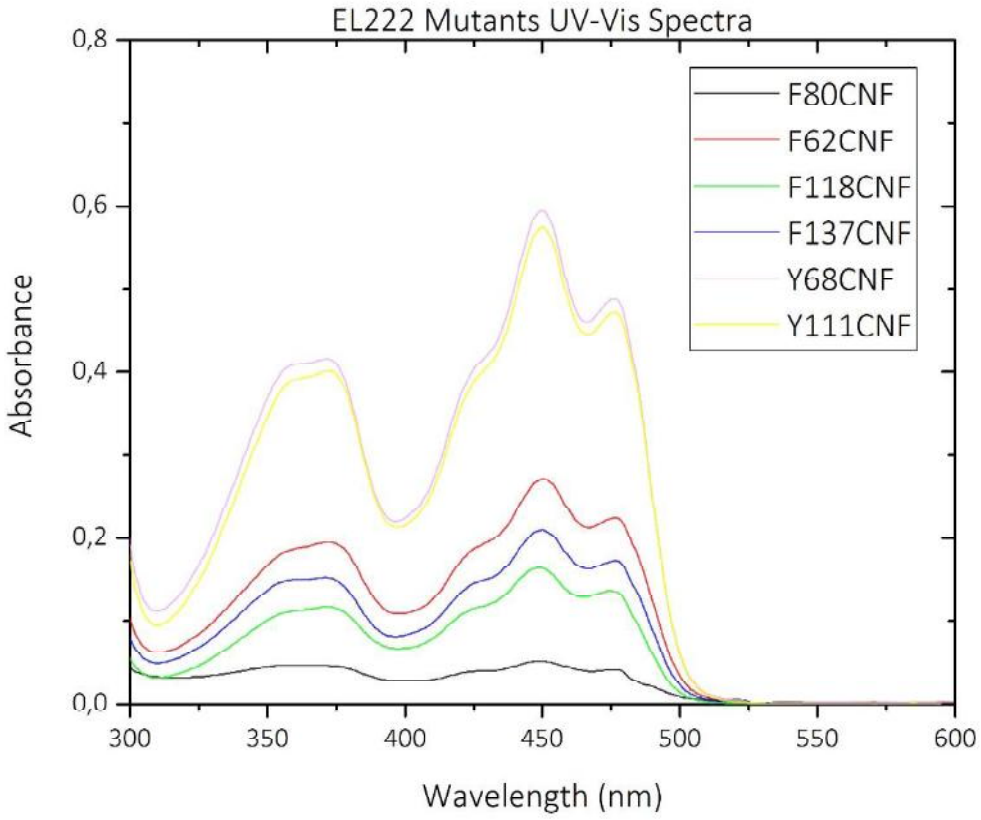


Figure 24 - UV-Vis spectra obtained for the six EL222 CNF mutants.

The spectra were compared and confirmed to be in line with results previously obtained and with results already obtained for EL222 wild type protein [5]. Due to this fact, it was possible to continue the experiments with these proteins.

4.3.2. FTIR spectroscopy.

In order to better quantify the mutant protein, the Direct Detect® Infrared Spectrometer (Merck) was used. This method only requires the use of 2 µL of sample and 2 µL of buffer.

Figure 25 shows the FTIR spectra obtained. There is a peak between 1600 cm^{-1} and 1700 cm^{-1} that corresponds to the Amide I vibration. Such band appears in proteins and peptides and arises mainly from the C=O stretching vibration.

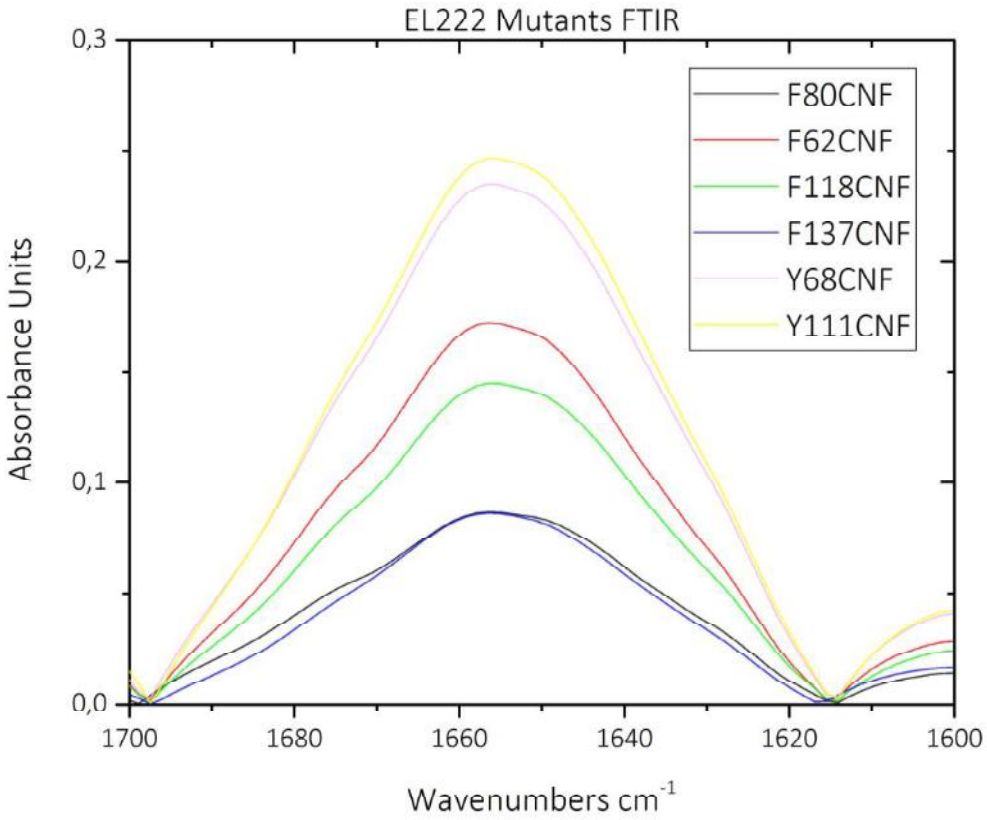


Figure 25 - FTIR spectra of the six EL222 protein mutants.

These analysies were repeated with other samples of the same batch of mutant proteins, but this time using a different FTIR equipment to obtain more complete, controlled and reliable results. In this next experiment it was identified the vibration from the cyano group (2220-2250 cm^{-1}) of the CNF mutation and to monitor alterations in the protein conformation in the passage from the dark state to the light state However, the results obtained are still being analysed and interpreted by the rest of the team.

4.4. Characterization of CarH proteins.

4.4.1. Mass spectrometry

In order to confirm that the expressed and purified protein was the correct one, an experiment was performed to obtain the molecular mass.

At the end of purification, one dark and one light state sample of CarH protein were collected and the concentration adjusted to 5 mg/ml. The samples were analyzed at CMS. Figure 26 shows the mass spectra obtained for CarH protein in the light state.

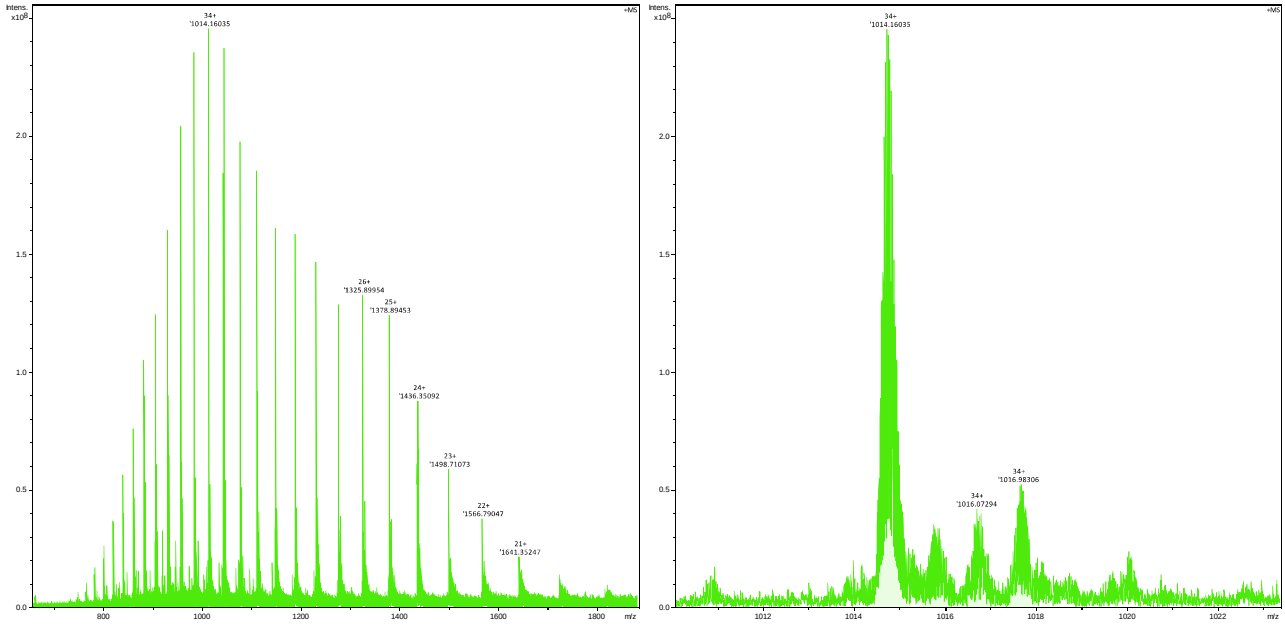


Figure 26 – Mass plots obtained for CarH full length protein.

The mass obtained from the experiment was 34449.22 g/mol and the expected one was 34450.98 g/mol. The expected value was obtained by summing the apo-protein mass in the light state (monomer) plus the mass of the cobalamin cofactor. The CarH protein sample was used in the light state as it proved to be the most stable sample. The spectra are very clear and with little noise, which shows a pure CarH protein containing cobalamin (i.e. AdoCbl lacking the 5'-adenosyl moiety) covalently bound.

By analyzing these results, it becomes possible to state that at the end of the expression and purification process, the correct protein was obtained in a very pure state.

4.4.2. UV-Vis spectroscopy

To determine the light-induced changes in the chromophore of CarH full-length protein and CarH cobalamin domain only, an experiment was performed to obtain the UV-Vis spectrum. For this purpose, the sample obtained from the purification processes was diluted 70-fold and its spectrum (dark state) was collected. Then the sample was irradiated with 34 mJ of green light for 10 seconds (or 60 seconds, in case of CarH cobalamin domain only) and the spectrum was collected again, this time being representative of the lightstate. In Figure 27 it is possible to compare the spectra of the same CarH sample in the dark state and light states.

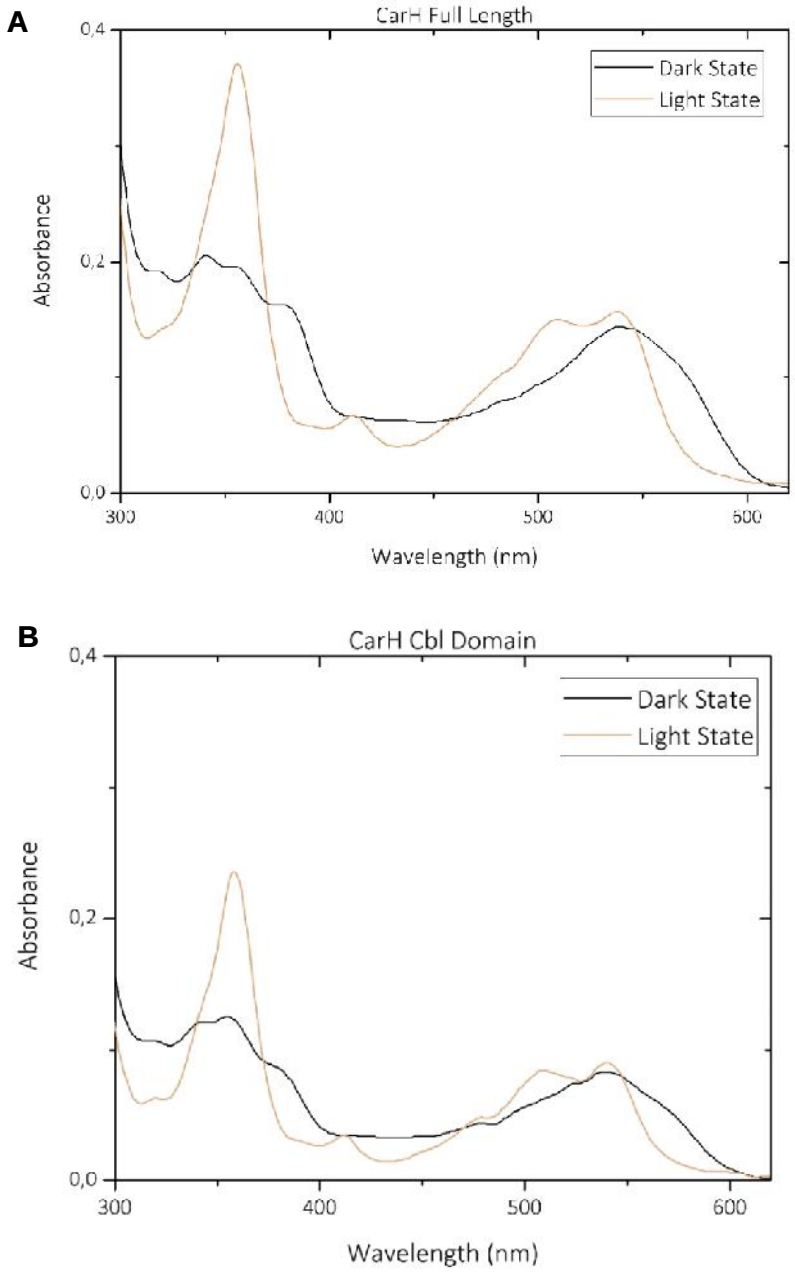


Figure 27 - Uv-Vis spectra for CarH protein showing dark state Vs light state. A) CarH full-Length, . B) CarH cbl domain only, green light. The light state spectra were obtained after irradiation with 3.4 mW of power and a wavelength of 530 nm.

By looking at Figure 27 and comparing the two spectra obtained, it is possible to state that there is a clear change in the cofactor structure when they pass from the dark state to the light state.

In order to verify protein sensitivity to light, the same assays were performed using different wavelengths (LEDs) at the same total energy of 34 mJ. Figure 28 shows the results obtained for the three LEDs tested.

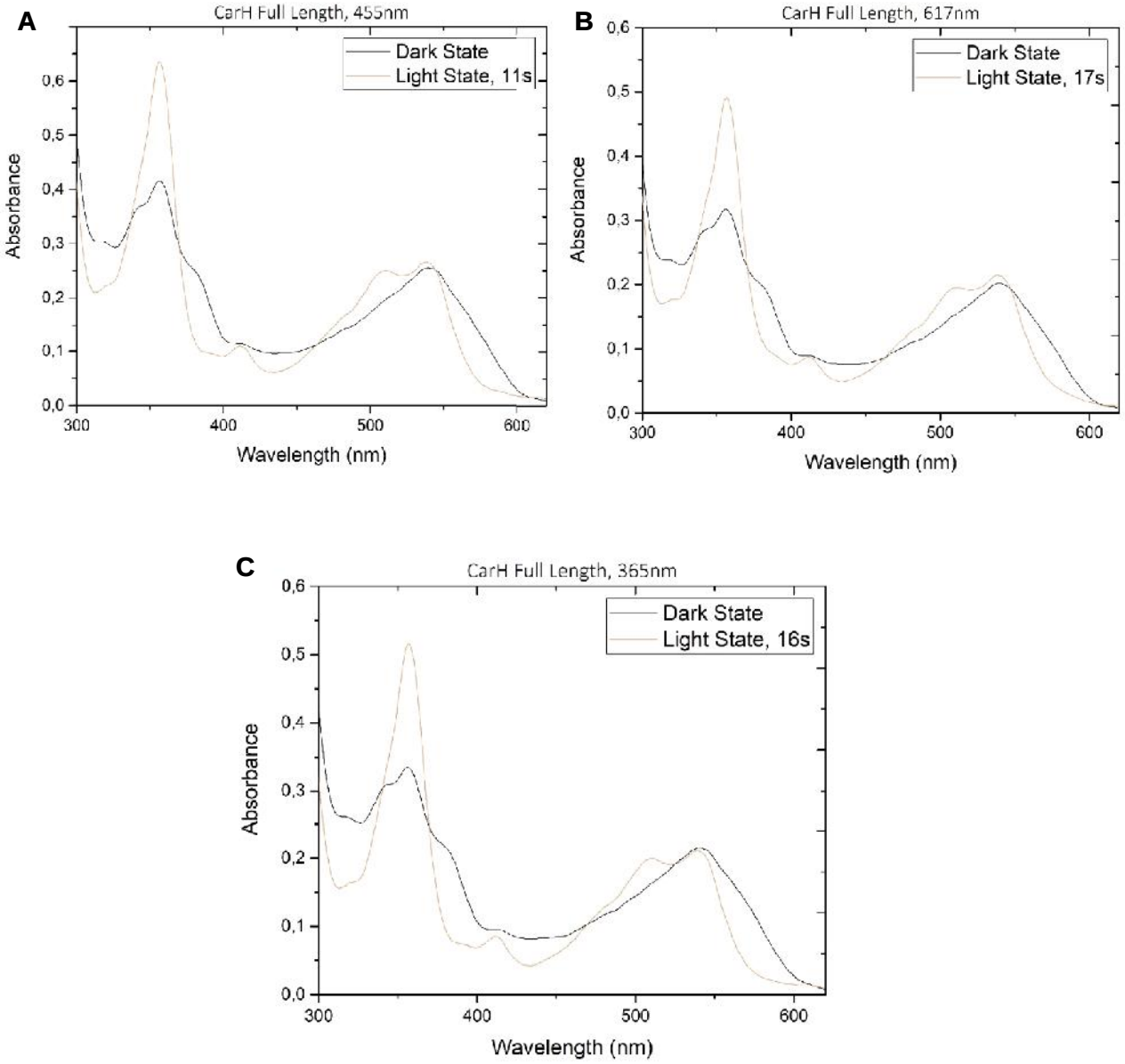


Figure 28 - UV-Vis spectra of CarH full-length protein dark vs light state. A) 455nm – blue light, B) 617nm –red light, C) 365nm – ultra-violet light.

As can be seen in Figure 28, all LEDs were able to activate the protein. This result confirms the light-sensitivity of the protein under study. It is surprising that red light (617 nm) can activate CarH. This may be due to a total energy sufficient to cleave the Co-C bond or to the relatively broad emission bands of the LEDs used that may leak into the green region of the spectra. The use of lower energies, optical filters and narrow-band emitters (e.g. laser diodes) may help understanding this issue.

4.4.3. MST

In order to study the interaction between protein and DNA, a microscale thermophoresis (MST) experiment was performed.

For this experiment, a constant concentration of 10 nM Cy5 DNA was used and the protein concentration was increased gradually. Figure 29 shows the graph obtained for the CarH protein MST experiment.

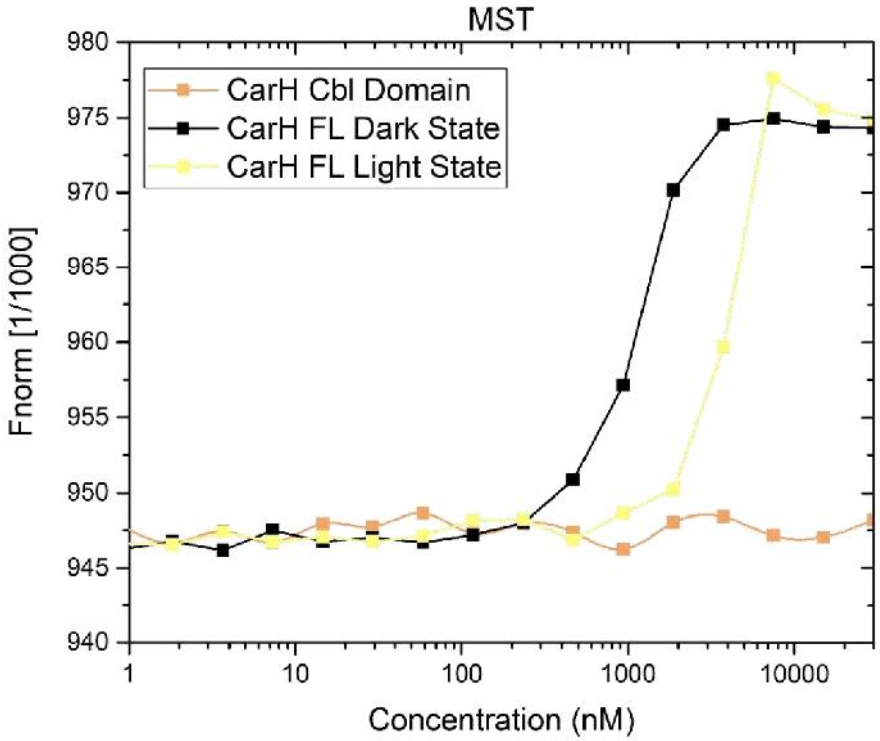


Figure 29 - Microscale thermophoresis of CarH protein. The DNA concentration was kept fixed at 10nM DNA-Cy5 and the protein concentration was varied as indicated in the figure.

As can be seen from Figure 29, the CarH full-length protein was first measured in the dark state, then the capillaries were irradiated and read again and the affinity for DNA binding was found to decrease, as expected. Finally, as a negative control, the CarH cobalamin domain only protein was measured and found not to interact with DNA, as expected, since this protein version lacks the domain for DNA binding.

The values for the dissociation constant (K_D) were calculated through the points obtained in the MST. Despite slight fluctuations in the calculated constants, one can safely

conclude that the affinity between CarH and DNA is higher in the light than in the dark state, as previously shown [31]. The values are shown in Table 16.

Table 16 - K_D values obtained by MST applied to CarH protein.

State of CarH Full-length	K_D (nM)
Dark	310
Light	5700

4.4.4. FTIR spectroscopy.

The last experiments were carried out with CarH proteins in order to know its secondary structure and light-induced conformational changes. The protein buffer was first exchanged into D_2O-based buffer to eliminate H_2O interference.

In figure 30 it is possible to observe the spectrum of CarH protein obtained with the transmission cell. The most important feature is the peak at the wavenumber of 1649.78 cm^{-1}, as this indicates the existence of the Amide I vibration which is due to the double bound between carbon and oxygen of amide groups, that are found mainly in peptide bonds.

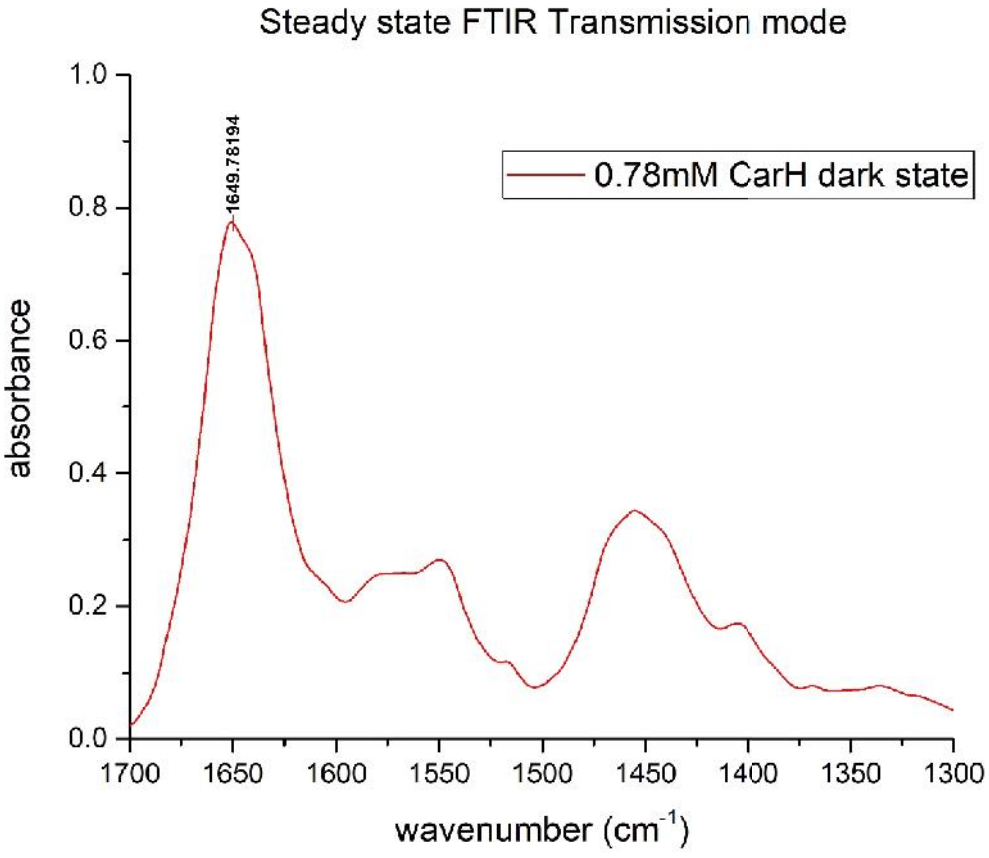


Figure 30 – FTIR spectrum of CarH full-length protein in the dark state.

Then the different spectra obtained in the dark state and the light state were compared. Figure 31 shows the spectra obtained. Predictably, in the dark vs dark (D-D) and light vs. light (L-L) spectrum, no major changes are observed as these two measurements also function as a control of the experiments. The light vs dark (L-D) spectrum shows the changes the protein undergoes upon illumination. In such spectrum, the most important information are the two positive peaks present between the 1600-1650 cm^{-1} wavenumbers, as they indicate the presence of α-helix and β-sheet. In other words, light-state CarH increases its secondary structure content (both α-helices and β-sheets) relative to the dark-state. This result is in agreement with the known crystallographic structures of CarH [31], suggesting similar conformational changes in both aqueous solutions and crystals. Further experiments with time-resolved FTIR will help to elucidate the sequence of folding events in the dark-to-light transition of CarH.

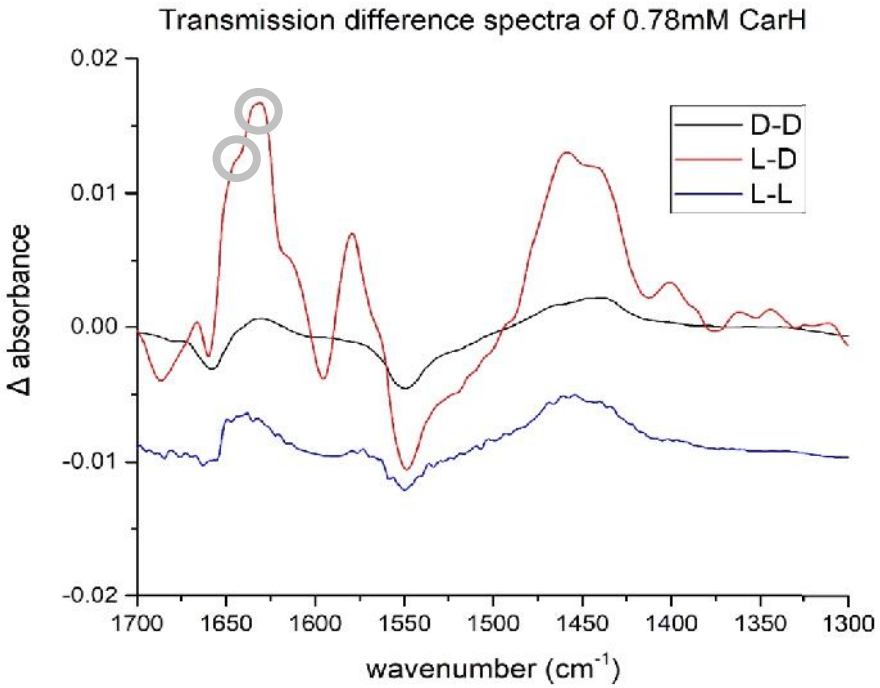


Figure 31 – Difference FTIR spectra of CarH full-length protein. Dark minus dark (D – D), light minus dark (L – D) and light minus light (L-L) spectra are shown.

4.4.5. Circular Dichroism (CD) spectroscopy

The last CarH experiment to obtain information on its secondary structure was the circular dichroism. For this experiment, a protein sample was given to the CMS facility and the results sent later. Figure 32 shows the resulting CD spectra.

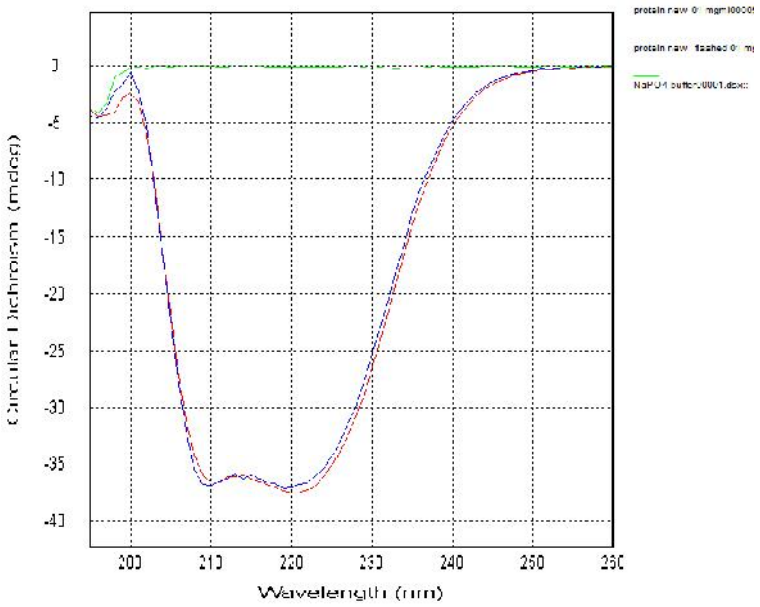


Figure 32 - CD spectra of CarH protein.

The protein was first measured in the dark state, irradiated and measured again in the light state. The percentages of secondary structure obtained for each of the states are shown in table 17.

Table 17 - Secondary structure content of CarH full-length proteins estimated from CD.

Structure	200-260 nm
Alpha-Helix	95.4 %
Antiparallel	0.4 %
Parallel	0.5 %
Beta-Turn	6.2 %
Random Coil	3.0 %
Total Sum	105.6 %

No major changes in secondary structure content are observed by CD suggesting that this technique is much less sensitive to conformational changes than FTIR spectroscopy.

5. CONCLUSIONS

With the techniques, technologies and expertise available at IBT (BIOCEV) it was possible to elaborate a realistic plan for the study of two DNA-binding photoreceptors: CarH and EL222 proteins.

The work plan for the 5-month internship was to better understand the structure and the pathway the proteins follow from dark to light states.

CarH is one of the most recently discovered photosensitive proteins and has many aspects to be unravelled. In the case of CarH protein, expression and purification were performed and optimized in a very positive manner, becoming at the end of the internship a simple protocol for the operator. Most experiments performed on this protein went as intended, however it is necessary to point out that there are some limitations in certain experiments. The experiment with infrared spectroscopy was the most challenging since there are no FTIR spectra published in articles to compare with our results.

By analysing all the experiments carried out on the CarH protein, it is possible to state that: i) The protein used in the studies is in a very pure form, ii) It is extremely light-sensitive, and iii) The secondary structure content (α-helix and β-sheets) is increased upon irradiation.

As for the EL222 protein, the wild type is already well characterized and the protein evaluation processes are already well optimized. However, it is necessary to increase our understanding of EL222 and LOV proteins in general. Non-canonical amino acids, like CNF may pave the way for site-specific and time-resolved studies of protein structural dynamics.

In conclusion, the entire internship carried out in this research centre over the 5 months was very satisfactory, and it was possible to make significant advances in the field of photosensory proteins, more specifically focusing on CarH and EL222 proteins.

The work done with these two proteins was far from complete. Importantly, all the experiments performed and conclusions drawn open the door for new questions and new points of views for study. With the continuation of the work left, it will be possible in the near future to make fundamental discoveries in the fields of protein folding and dynamics, as well as more applied advances in the fields of synthetic biology and optogenetics.

6. BIBLIOGRAPHY

[1] https://www.biocev.eu/en, *consulted on March 8, 2019*

[2] http://www.ibt.cas.cz/index.html, *consulted on November 26, 2019*

[3] Herrou, J.; Crosson, S. *Function, structure, and mechanism on bacterial photosensory LOV proteins*. Nat Rev Microbiol, 2012, 713-723, Chicago

[4] Kottke, T.; Xie, A.; Larsen, D.S.; Hoff, W.D. *Photoreceptors Take Charge: Emerging Principles for Light Sensing*. Annual Review of Biophysics, 2018, 291-313

[5] Iuliano, J.A.; Gil, A.A.; Laptenok, S.P. *Variation in LOV Photoreceptor Activation Dynamics Probed by Time-Resolved Infrared Spectroscopy*. Biochemistry, 2018, 620-630

[6] Ortiz-Guerrero, J.M.; Polanco, M.C.; Murillo, F.J., *Light-dependent gene regulation by a coenzyme B12-based photoreceptor.* Proc Natl Acad Sci USA, 2011

[7] Padmanabhan, S.; Castano-Pérez, R.; Elías-Arnanz, M. *B12-based photoreceptors: from structure and function to applications in optogenetics and synthetic biology.* Current Opinion in Structural Biology, 2019, 47-55

[8] Padmanabhan, S.; Jost, M.; Drennan, C.L.; Elías-Arnanz, M. *A new facet of vitamin B12: gene regulation by cobalamin-based photoreceptors.* Annu Rev Biochem, 2017, 485-514

[9] Jones, A.R. *The photochemistry and photobiology of vitamin B12.* Photochem Photobiol Sci, 2017, 820-834

[10] https://www.uniprot.org/uniprot/Q2NB98, *consulted on November 27, 2019*

[11] https://www.uniprot.org/uniprot/Q53W62, *consulted on November 27, 2019*

[12] Nash, A.I.; McNulty, R.; Shillito, M.E. *Structural Basis of photosensitivity in a bacterial light-oxygen-voltage/helix-turn-helix (LOV-HTH) DNA-binding protein*. Proc Natl Acad Sci USA, 2011

[13] Zoltwski, B.D.; Motta-Mena, L.B.; Gardner, K.H. *Blue light-induced dimerization of a bacterial LOV-HTH DNA-binding protein*. Biochemistry, 2013

[14] Takakado, A.; Nakasone, Y.; Terazima, M. *Photoinduced dimerization of photosensory DNA-binding protein EL222 and its LOV domain.* Phys Chem Chem Phys, 2017

[15] Rivera-cancel, G.; Motta-Mena, L.B.; Gardner, K.H. *Identification of natural and artificial DNA substrates for light-activated LOV-HTH transcription factor EL222*. Biochemistry, 2012

[16] Takakado, A.; Nakasone, Y.; Terazima, M. *Sequential DNA Binding and Dimerization Processes of the Photosensory Protein EL222*. Biochemistry, 2018

[17] https://app.biorender.com/signin, *consulted on November 1, 2019 (used for drafting the schemes)*

[18] Young, D.D.; Schultz, P.G. *Playing with the Molecules od Life.* ACS Chem Biol, 2018

[19] Schultz, K.C.; Supekova, L.; Ryu, Y. *A genetically encoded infrared probe.* J Am Chem Soc, 2006

[20] Pi, J.; Seal, L. *Mass Spectrometry Coupled Experiments and Protein Structure Modeling Methods.* International Journal of Molecular Sciences, 2013, 20635-20657

[21] Gingras, A.; Gstaiger, M.; Raught, B.; Aebersold, R. *Analysis of protein complexes using mass spectrometry.* Nature Reviews, volume 8, 2007, 645-654

[22] Amenabar, I.; Poly, S.; Nuansing, W.; Hubrich, E.H. *Structural analysis and mapping of individual protein complexes by infrared nanospectroscopy.* Nature Communications, 2013

[23] Kottke, T.; Lórenz-Fonfría, V.A.; Henerle, J. *The Grateful Infrared: Sequential Protein Structural Changes Resolved by Infrared Difference Spectroscopy.* J Phys Chem B, 2017

[24] Seidel, S.A.I.; Dijkman; P.M.; Lea, W.A. *Microscale Thermophoresis Quantifies Biomolecular Interactions under Previously Challenging Conditions.* Methods, 2013, 301-315

[25] Jerabek-Willemsen, M.; André, T.; Wanner, R. *MicrosScale Thermophoresis: Interaction analysis and beyond.* Journal of Molecular Structure, 2014, 101-113

[26] Greenfild N.J. *Using circular dichroism spectra to estimate protein secondary structure.* Nat Protec, 2006, 2876-2890

[27] https://international.neb.com/products/c2987-neb-5-alpha-competent-e-coli-high-efficiency#Product%20Information, *consulted on November 20, 2019*

[28] https://international.neb.com/products/c2527-bl21de3-competent-e-coli#Product%20Information_Advantages%20and%20Features, *consulted on November 20, 2019*

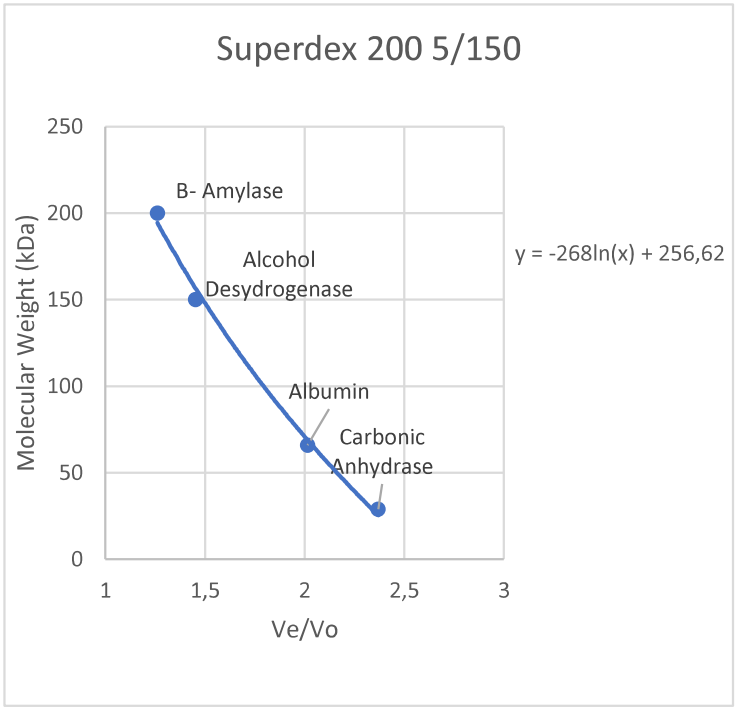
[29] https://international.neb.com/protocols/2013/12/13/pcr-using-q5-high-fidelity-dna-polymerase-m0491, *consulted on April 10, 2019*

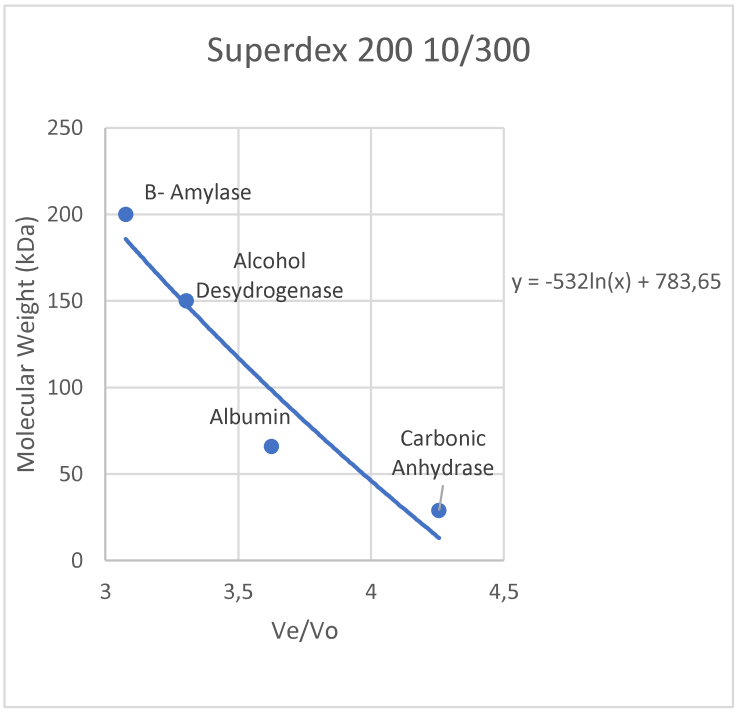
[30] http://tools.thermofisher.com/content/sfs/brochures/110-dna-insert-ligation-sticky-end-and-blunt-end-into-vector-dna.pdf, *consulted on April 10, 2019*

[31] Jost, M.; Fernández-Zapata, J.; Polanco, M.C. *Structural basis for gene regulation by a B12-dependent photoreceptor.* Nature, 2015

7. ANNEX

7.1. Calibration of size-exclusion columns

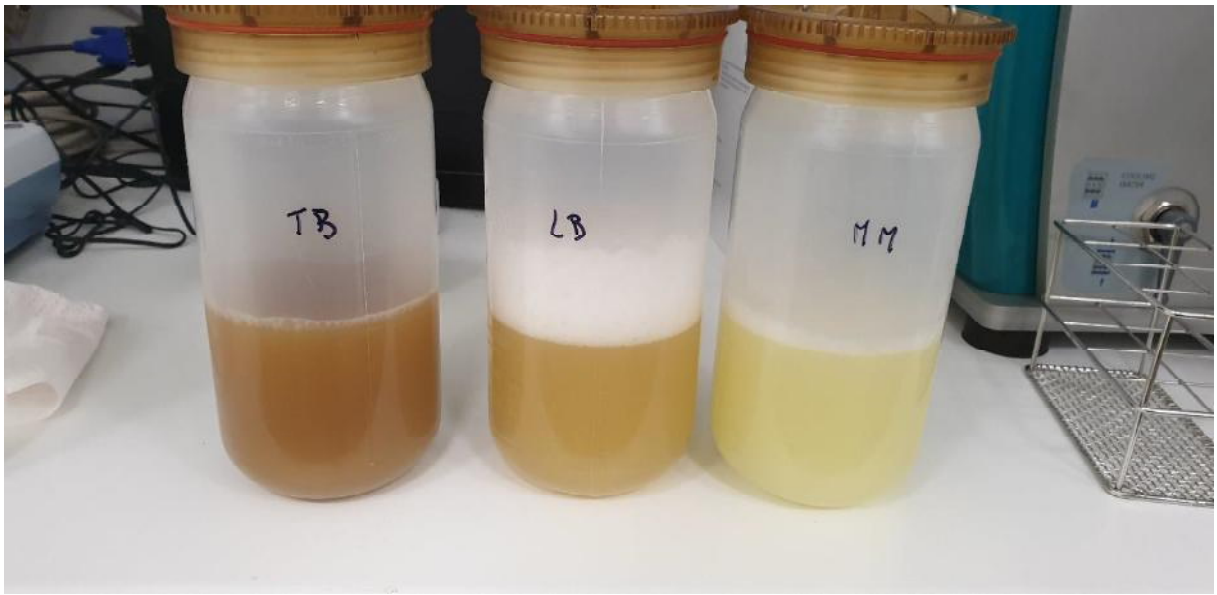




7.2. Images of the protein expression and purification process

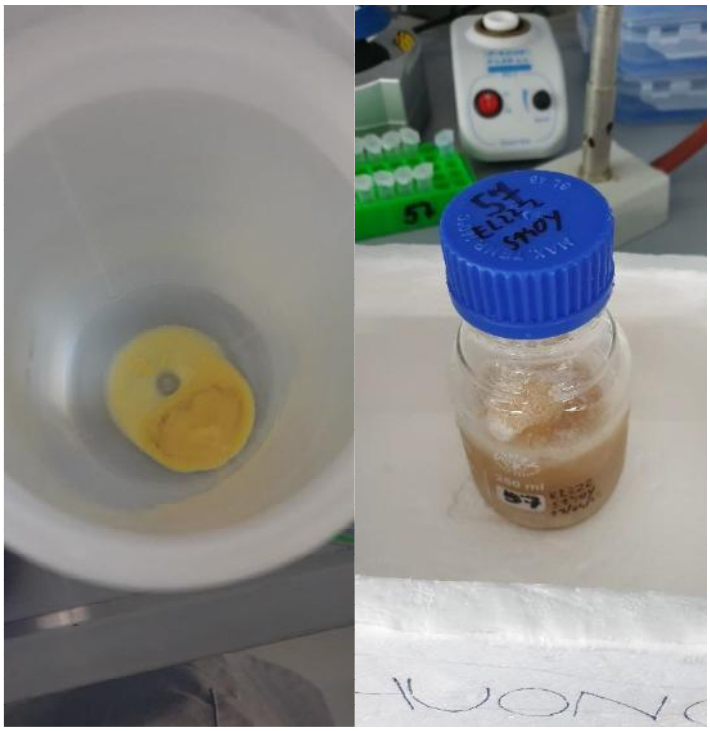
- Induction of cell expression


- Culture of EL222 Mutants

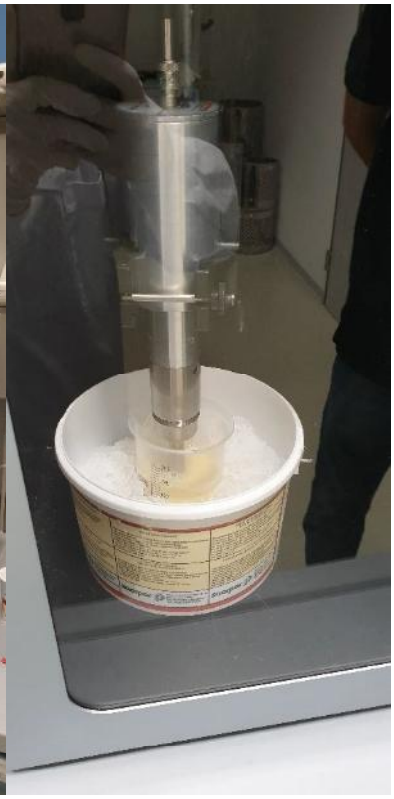


- Centrifugation for BL21(DE3) cell collection

- BL21(DE3) Cell Pellet

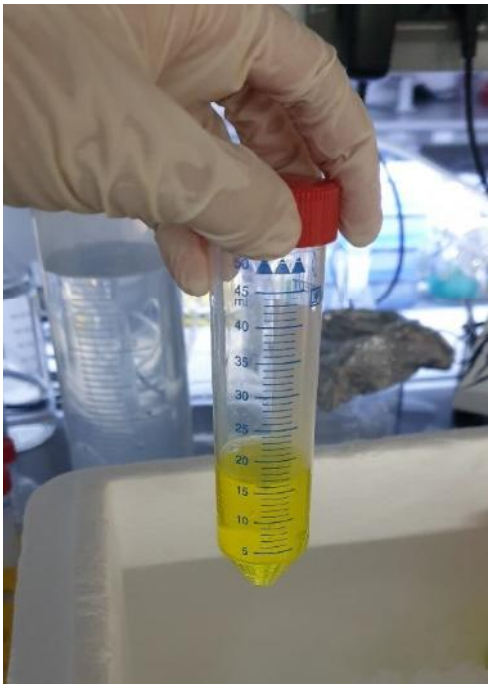


- Sonication

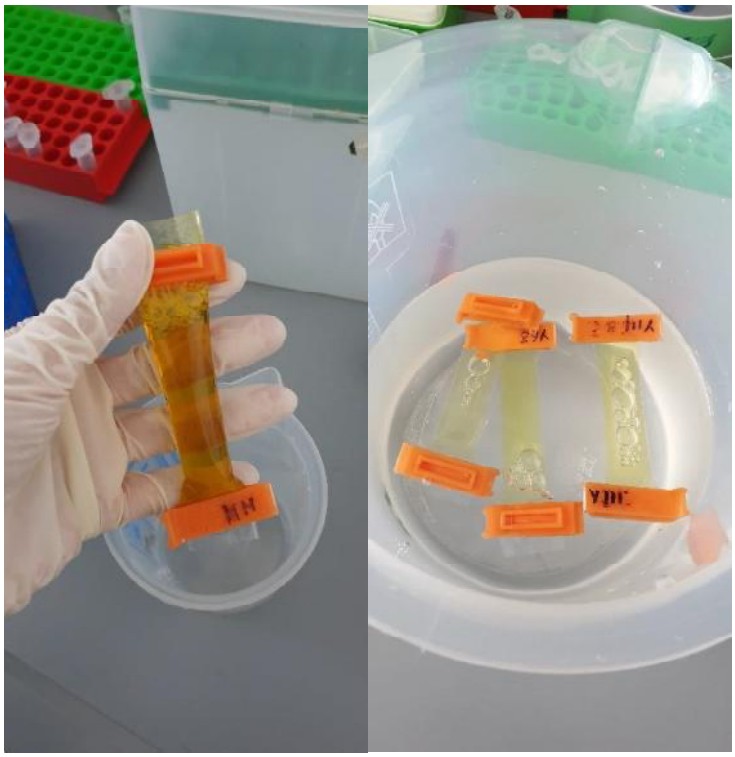
- Microfluidizer

- Protein EL222 wild type for purification



- 1st Affinity chromatography

- Dialysis (EL222)



- Size-exclusion Chromatography and ion-exchange Chromatography (Protein EL222)

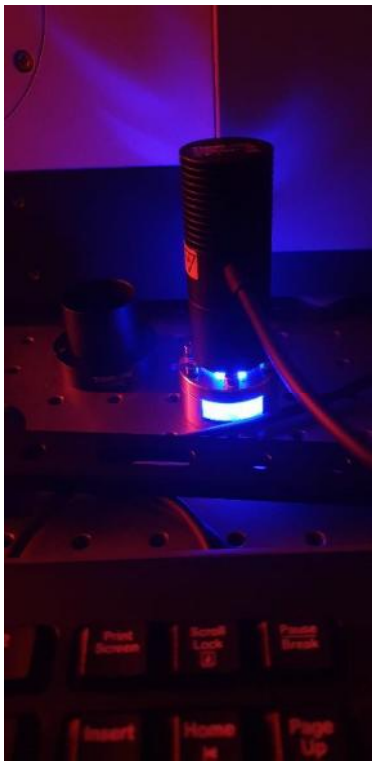
- SDS-PAGE equipment

- Liquid nitrogen to freeze purified proteins

- Purified proteins (EL222, CarH)

- FTIR sample preparation

- Irradiation of samples for FTIR



7.3. Materials and equipment used

- MicroPulser Electroporator

- Macherey-Nagel™ NucleoSpin™ Gel and PCR Clean-up Kit

- Thermo Scientific™ MaxQ™ 5000 Floor-Model Shakers

- Thermo Scientific™ MaxQ™ 4000 Benchtop Orbital Shakers

- Avanti J-30I

- ProFlex™ 3 x 32-well PCR System

- Qsonica Sonicator Q700

- EmulsiFlex-C3

- Rocker 35 EZ Large Capacity Lab Rocker

- Spectrum™ Spectra/Por™ 7 Membrane Tubing (Sulfur and Heavy Metal Free): 6000 to 8,000 Dalton MWCO

- NGC Discover™ 100 Chromatography System #7880010

- VERTEX 70v FTIR spectrometer (left) and laser (right)

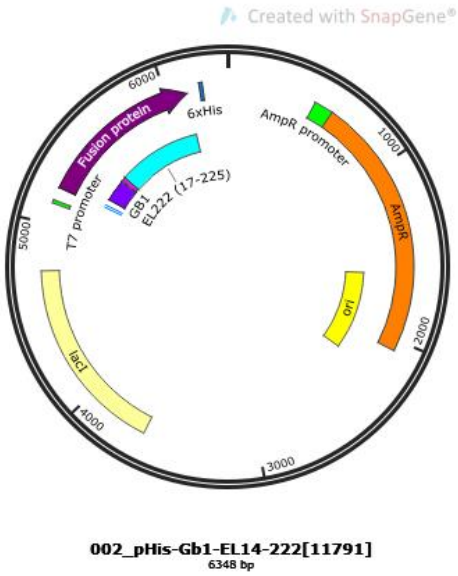
- Monolith NT.115

- Direct Detect® Infrared Spectrometer

- LED’s Thorlabs

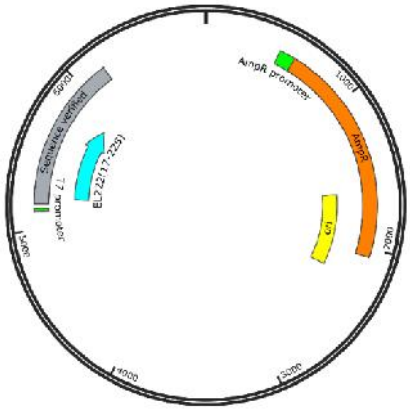
7.4. Plasmids

- EL222 Wild type Plasmid #02



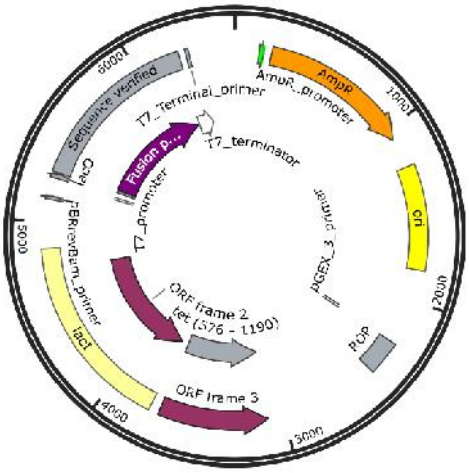
002_pHis-Gb1-EL14-222[11791]
6348 bp

- EL222 Y68TAG Plasmid #106



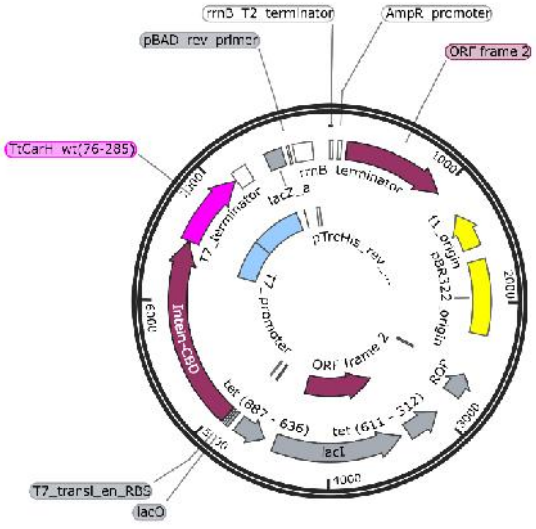
106_pEL14-222-Intein-CBD-12His_G2A_Y68TAG[11789]
6876 bp

- CarH full-Length Plasmid #60



060_pET15br_TtCarH[11788]
6558 bp

- CarH cbl Domain only Plasmid #59



059_pTYB12_TtCarH(76-285)[11790]
7980 bp