Agradecimentos

Ao finalizar este trabalho que faz parte de uma das mais importantes etapas que já passei, queria desde já agradecer as pessoas que me permitiram desenvolver e concluir este trabalho e que me acompanharam neste caminho.

Agradeço em primeiro lugar ao Doutor Pedro Castanheira, ao Professor Doutor Euclides Pires e ao Professor Doutor Carlos Faro pelas óptimas condições proporcionadas bem como o apoio na realização deste trabalho de investigação para obtenção de grau de Mestre.

Ao Doutor Pedro Castanheira um forte agradecimento pela orientação científica proporcionada durante este percurso.

A todas as pessoas da Unidade de Biotecnologia Molecular, à Carla, à Professora Isaura pelas sugestões apoiantes de alternativas para o trabalho, pela boa disposição e disponibilidade. À Marisa, ao André, à Ana Rita, à Ana Sofia, à Liliana, ao Pedro Curto ao Rui e a Joana um obrigado pelas dicas ajudas e conselhos dados, pelo bom e descontraído ambiente vivido, inclusive os almoços e lanches, das conversas e momentos animados. Um obrigado também à D. Alda e à Ana.

Um obrigado extra ao Rui por toda a disponibilidade e por o que me ensinou e juntamente à Joana por ambos me ouvirem e me darem força naquele momento mais apertado.

Um forte obrigado aos meus pais por todo o apoio que me deram do início ao fim, por me ouvirem quando mais precisei, pelo vosso esforço, amor, dedicação, paciência e por acreditarem em mim. São e serão um orgulho para mim. À minha irmã Sofia e ao Miguel pela vossa amizade, preocupação e carinho.

Ao Mário por tudo o que acrescentas. Pelo apoio principalmente nos momentos decisivos, carinho animo e força e acima de tudo por nunca duvidares de mim. Ao meu pequeno sobrinho Henrique que tanto me animou e me fez descontrair nos bocadinhos que podia.

Às minhas amigas Sousa, Catarina e Cláudia, especialmente por serem as minhas companheiras este ano e termos partilhado as mesmas experiências.

A todos os meus colegas de Bioquímica que estiveram comigo nos últimos 5 anos e os bons momentos que tivemos.

Table of Contents

Abbreviations

AMPK: AMP-activated protein kinase

E. coli: Escherichia coli

EDTA: Ethylenediaminetetraacetic acid

GS: glycogen synthase

GSK3: glycogen synthase kinase 3

GST: Glutathione-S-transferase

HPLC: high pressure liquid chromatography

IMAC: Immobilized Metal ion Affinity Chromatography

IPTG: Isopropyl β-D-1-thiogalactopyranoside

LB: Luria Bertani Broth

MBP: Maltose Binding Protein

Nus-A: N-utilizing substance A

OD600: Optical Density, measured at a wavelength of 600nm

PCR: polymerase chain reaction

PKA: Protein *kinase* A

PP1: protein phosphatase 1

PVDF: Polyvinylidene Fluoride

SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SEC: Size-exclusion chromatography

TBS: Tris-Buffered Saline

TBS-T: Tris-Buffered Saline with Tween20

TEV: Tobacco Etch Virus

Trx: thioredoxin

UDP-glucose: uridine diphosphate glucose

WB: Western Blot

Abstract

Glycogen is a branched and soluble polymer of glucose, present in cells functioning as energy storage. Its accumulation and utilization, synthesis and degradation are controlled primarily by covalent phosphorylation and allosteric ligand binding of the enzymes involved in glycogen metabolism. It is suggested that such enzymes are controlled by two enzymes, laforin and malin, which form a complex that is able to ubiquitinate those targets and target them for degradation, as a way of controlling glycogen synthesis. Mutations in the genes coding for these two proteins are the cause of Lafora disease, a type of myoclonic progressive epilepsy. Laforin is a unique human dual-specificity phosphatase (DSP), as it contains a carbohydrate binding module (CBM) at its N-terminal along with the Dual-Specificity Phosphatase Domain (DSPD) at its C-terminal, whereas malin is an E3 ubiquitin ligase of RING type. Malin was shown to interact with laforin within the region between the CBM and DSPD domains. Based on this, the aim of this work is the optimization of expression and purification of protein to obtain sufficient amounts of protein for further identification of a minimal laforin amino acid sequence required for the interaction with malin.

To achieve this, several constructs coding for either the CBM or DSPD domains with increasing length of the inter-domain region were obtained by molecular biology techniques. The protein expression was performed using E. coli as host system, and the results have shown that, apart from the shorter CBM constructs, most protein is expressed in the form of insoluble inclusion bodies. The attempts to refold such proteins using the protocol previously developed for the refolding of laforin and its CBM proved unsuccessful. Laforin CBM domain was successfully expressed in soluble form and has been purified with a yield of 1,73 mg/L of expression media. Malin expression revealed that the protein is very poorly expressed in E. coli, being only detected by western-blot. Most of the protein was expressed as inclusion bodies, and despite some attempts to obtain soluble protein expression, no protein could be purified.

This work has evidenced the difficulties of expressing protein domains and the importance of choosing the correct boundaries of such domains. Nevertheless, this work paves the way for the further optimization necessary for the expression and purification of both malin and the laforin constructs developed during this work.

Resumo

O glicogénio consiste num polímero de glucose, solúvel e ramificado que se encontra presente no citoplasma celular e funciona como reserva energética. A acumulação e utilização, bem como a síntese e degradação são processos controlados pela regulação da atividade das enzimas envolvidas no metabolismo do glicogénio, por reações de fosforilação e interações alostéricas. Duas enzimas, a laforina e a malina, estão envolvidas nesta regulação enzimática. Formam um complexo que promove a ubiquitinação dos respectivos substratos e o seu direcionamento para degradação permitindo uma regulação da síntese do glicogénio. Mutações nos genes que codificam ambas as proteínas levam ao aparecimento da doença da Lafora que consiste num tipo de epilepsia mioclónica progressiva. A laforina é a única fosfatase de dupla especificidade (DSP) humana devido à presença de um respectivo domínio DSP no C-terminal apresentando também um módulo de ligação a carbohidratos (CBM) no N-terminal. Por sua vez, a malina é uma ligase de ubiquitina E3 do tipo RING finger e interage com a laforina na região entre os domínios CBM e DSPD. Com base nesta informação, o objectivo do projecto consistiu na optimização da expressão e purificação das proteínas de modo a obter quantidades suficientes para uma posterior identificação da região mínima da laforina necessária para interagir com a malina. Para isto, várias construções contendo quer o domínio CBM ou o domínio DSP com um aumento da sequência da região entre os domínios foram obtidas por técnicas de biologia molecular. A expressão de proteína foi feita utilizando a E. coli como sistema de expressão e de acordo com os resultados obtidos, à parte das construções mais curtas contendo o domínio CBM, a maior parte da proteína das restantes construções é expressa na forma de corpos de inclusão. Tentativas de refolding destas proteínas com base num protocolo previamente desenvolvido para expressão de laforina e do seu domínio CBM falharam. No entanto, o domínio CBM da laforina foi expresso na forma solúvel e purificado com um rendimento de 1,73 mg/L de meio de expressão. Relativamente à malina, os rendimentos de expressão em E. Coli são muito baixos no qual a proteína só é detectada em western-blot. A maior parte da proteína é expressa como corpos de inclusão e apesar de algumas tentativas no sentido de obter proteína solúvel, ainda assim não foi possível purificar a malina.

Este trabalho mostra a dificuldade na expressão de domínios proteicos e na importância de uma escolha correcta dos limites destes domínios. No entanto, este

trabalho abre caminho para futuras optimizações necessárias para a expressão e purificação da malina assim como das construções derivadas da laforina desenvolvidas ao longo deste trabalho.

1. Introduction

1.1. Glycogen function, localization and structure

Glycogen consists of a branched glucose polymer present in all mammals, stored mainly in skeletal muscle and liver. However this particle can also be synthesized and accumulated in the brain, kidneys, cardiac and smooth muscle, and in adipose tissue. This polymer functions as storage of energy and carbon and is used when blood glucose levels are low. The glucose levels in the cell are tightly regulated by two hormones - one is insulin, a hormone involved in cell signaling for glucose uptake and storage as glycogen when blood glucose levels are high; the other is glucagon, which is more secreted than insulin leading to glucose production from glycogen in the liver that is then sent to the bloodstream and distributed to the other cells. However, glucose released from glycogen in skeletal muscle is only used by this organ when rapid supply of energy is needed.

Relatively to the structure, glycogen is formed by primarily glucose chains with branching points with a certain frequency that will give the final and characteristic spherical and symmetric structure (fig. 1). This particle is localized in the cell cytoplasm due to its solubility characteristic conferred by the hydroxyl groups from the glucose monomers[1].

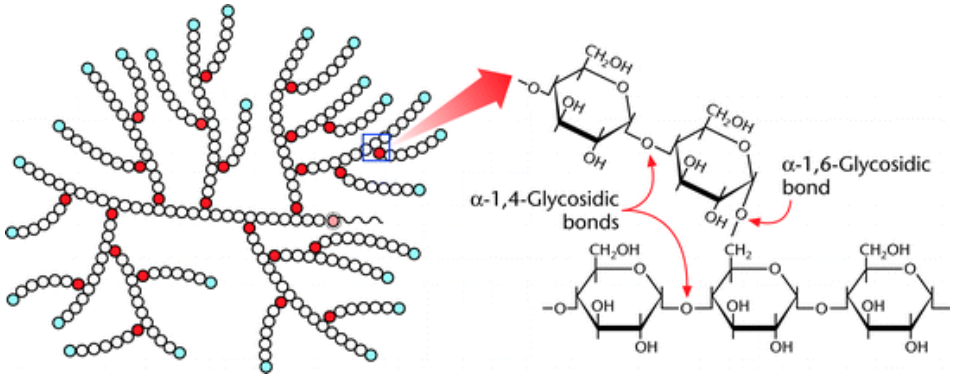

Figure 1. Structure of glycogen. Adapted from Marks DB, Marks AD and Marks CM (eds) (1996) Medical Biochemistry: A Clinical Approach.

1.2. Glycogen metabolism

1.2.1 Glycogen synthesis and degradation

The synthesis of glycogen is dependent on three enzymes. The first one is glycogenin (E.C. 2.4.1.186), a glycosyltransferase that initiates the synthesis by attaching 10 glucose units, forming a base chain as a starting point for the central synthesis by glycogen synthase (EC 2.4.1.11) and the branching enzyme (EC 2.4.1.18). Glycogen synthase is also a glycosyltransferase which catalyzes α-1-4-glycosidic linkages by adding a UDP-glucose monomer (Uridine diphosphate glucose), to ensure the elongation of the polyglucose chains. Finally the branching enzyme is responsible for introducing branching points. It cleaves an α-1-4 glycosidic linkage and excises a segment of an oligosaccharide forming an α-1-6 glycosidic linkage. Glycogen contains small amounts of phosphate covalently linked in form of C2 and C3 phosphomonoesters that are incorporated by glycogen synthase through β-phosphate UDP-glucose substrate. For instance in muscle glycogen it is estimated the presence of one phosphate per 600-1500 glucose residues. This event is a side reaction of GS, at a rate of one phosphate per 10 000 glucoses, which is believed to be a catalytic error[2].

The degradation of glycogen involves two enzymes. Glycogen phosphorylase which cleaves α-1-4 glycosidic linkages to yield glucose-1-phosphate which is then converted to glucose-6-phosphate. Besides, the debranching enzyme (amylo-1,6-glucosidase,4-α-glucanotransferase) (EC 3.2.1.68) is responsible for cleavage of α-1-6 glycosidic linkage.

1.2.2 Glycogen metabolism regulation

Glycogen homeostasis is maintained by the regulation of synthesis and degradation, through mechanisms involving phosphorylation and even allosteric regulation[3,4]. At the level of synthesis, glycogen synthase is inhibited by phosphorylation through kinases such as AMPK, PKA, and GSK3. On the other hand, GS is only activated by the type-1 protein phosphatase (PP1). The PP1 can also inactivate glycogen phosphorylase leading to glycogen accumulation. The action and recruitment of PP1 to glycogen is mediated by the glycogen targeting protein R5/PTG, which functions as a scaffold since

it binds to PP1, glycogen and GS or GP, targeting PP1 to dephosphorylate each of the two enzymes, activating GS and inactivating GP.

1.2.3 The involvement of laforin and malin proteins in glycogen synthesis regulation

1.2.3.1 Laforin and malin characterization

Laforin is a human protein phosphatase encoded by the EPM2A gene with 3128 bp and four exons forming the 996 bp coding sequence (fig. 2).This gene is located on chromosome 6q24. It is expressed mainly in liver, heart, skeletal muscle, pancreas and brain[5,6]. Laforin has 331 amino acid residues and is present in vivo in monomeric and/or dimeric forms depending on cell redox conditions[7].

```
1    atgcgcttccgctttggggtggtggtgccacccgccgtggccggcgcccggccggagctg
1     M  R  F  R  F  G  V  V  V  P  P  A  V  A  G  A  R  P  E  L
61   ctggtggtggggtcgcggcccgagctggggcgttgggagccgcgcggtgccgtccgcctg
61    L  V  V  G  S  R  P  E  L  G  R  W  E  P  R  G  A  V  R  L
121  aggccggccggcaccgcggcgggcgacggggccctggcgctgcaggagccgggcctgtgg
41    R  P  A  G  T  A  A  G  D  G  A  L  A  L  Q  E  P  G  L  W
181  ctcggggaggtggagctggcggccgaggaggcggcgcaggacggggcggagccgggccgc
61    L  G  E  V  E  L  A  A  E  E  A  A  Q  D  G  A  E  P  G  R
241  gtggacacgttctggtacaagttcctgaagcgggagccgggaggagagctctcctgggaa
81    V  D  T  F  W  Y  K  F  L  K  R  E  P  G  G  E  L  S  W  E
301  ggcaatggacctcatcatgaccgttgctgtacttacaatgaaaacaacttggtggatggt
101   G  N  G  P  H  H  D  R  C  C  T  Y  N  E  N  N  L  V  D  G
361  gtgtattgtctcccaataggacactggattgaggccactgggcacaccaatgaaatgaag
121   V  Y  C  L  P  I  G  H  W  I  E  A  T  G  H  T  N  E  M  K
421  cacacaacagacttctattttaatattgcaggccaccaagccatgcattattcaagaatt
141   H  T  T  D  F  Y  F  N  I  A  G  H  Q  A  M  H  Y  S  R  I
481  ctaccaaatatctggctgggtagctgccctcgtcaggtggaacatgtaaccatcaaactg
161   L  P  N  I  W  L  G  S  C  P  R  Q  V  E  H  V  T  I  K  L
541  aagcatgaattggggattacagctgtaatgaatttccagactgaatgggatattgtacag
181   K  H  E  L  G  I  T  A  V  M  N  F  Q  T  E  W  D  I  V  Q
601  aattcctcaggctgtaaccgctacccagagcccatgactccagacactatgattaaacta
201   N  S  S  G  C  N  R  Y  P  E  P  M  T  P  D  T  M  I  K  L
661  tatagggaagaaggcttggcctacatctggatgccaacaccagatatgagcaccgaaggc
221   Y  R  E  E  G  L  A  Y  I  W  M  P  T  P  D  M  S  T  E  G
721  cgagtacagatgctgccccaggcggtgtgcctgctgcatgcgctgctggagaagggacac
241   R  V  Q  M  L  P  Q  A  V  C  L  L  H  A  L  L  E  K  G  H
781  atcgtgtacgtgcactgcaacgctggggtgggccgctccaccgcggctgtctgcggctgg
261   I  V  Y  V  H  C  N  A  G  V  G  R  S  T  A  A  V  C  G  W
841  ctccagtatgtgatgggctggaatctgaggaaggtgcagtatttcctcatggccaagagg
281   L  Q  Y  V  M  G  W  N  L  R  K  V  Q  Y  F  L  M  A  K  R
901  ccggctgtctacattgacgaagaggccttggcccgggcacaagaagatttttccagaaa
301   P  A  V  Y  I  D  E  E  A  L  A  R  A  Q  E  D  F  F  Q  K
961  tttgggaaggttcgttcttctgtgtgtagcctgtag
321   F  G  K  V  R  S  S  V  C  S  L  -
```

Figure 2. Human laforin coding sequence and amino acid sequence (UniProt: O95278). The nucleotide sequence is in black and the amino acid sequence is in grey, while the CBM domain is underlined and the DSP domain is dashed underlined.

Two isoforms are produced by alternative splicing from EPM2A gene and both are similar from residues 1 to 309, however they have different C-terminal. The canonical form is isoform laforin 1-331 which localizes in the cytoplasm and endoplasmic reticulum. The minor isoform laforin 1-317 is targeted to the nucleus and lacks phosphatase activity[8,9]. Laforin contains a carbohydrate-binding module (CBM) from residues 1 to 124, at the N terminal and a C-terminal dual-specificity phosphatase domain (DSPD) (residues 157 to 326). The laforin CBM belongs to the family CBM20, one of the 64 families of CBMs in which classification was based on evolutionary relationships, polypeptide folds and substrates (CAZY database). The structure of CBM 20 family consists of seven β-strands forming an open-sided distorted β-barrel (fig. 3). The CBM allows laforin to bind to complex carbohydrates such as glycogen and starch, amylopectin and amylose, and contains three conserved aromatic residues involved in this binding[10]. Since no crystal structure of CBM is yet available, a model of CBM is suggested based on the most similar structure used as template (fig. 3).

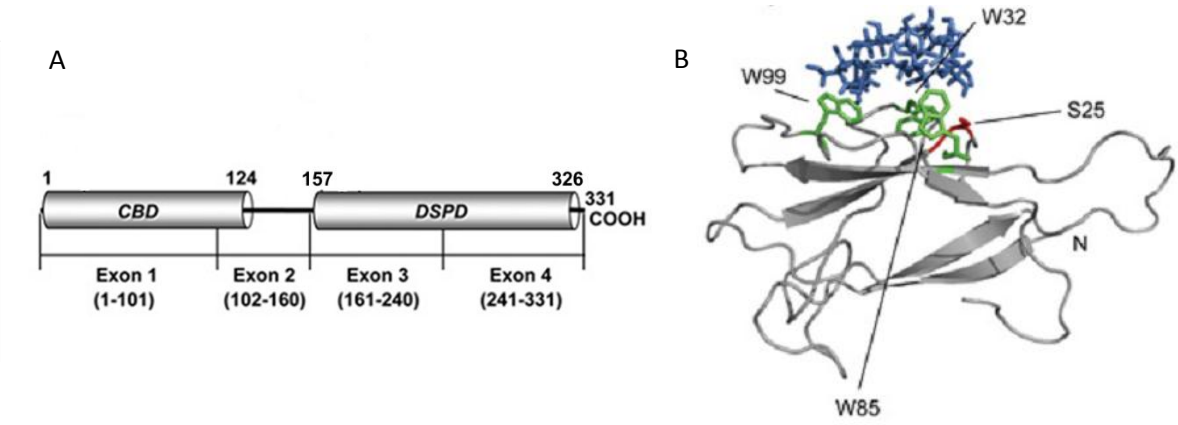

Figure 3. Structure model of Laforin CBM. |A. Schematic representation of human laforin where the CBM and DSPD domains are evidenced. **|B.** The CBM model was generated in *silico* using the crystal structure of *Geobacillus stearothermophilus* cyclodextrin glycosyltransferase as template (PDB: 1CYG). The homology model suggests that laforin CBM folding consists of characteristic two β-sheets fold, each consisting of three to six antiparallel β-strands and with the N- and C-termini pointing towards opposite ends. Conserved aromatic residues involved in carbohydrate binding are outlined in green: Trp^{32} Trp^{85}, and Trp^{99}. Adapted from Roma-Mateo et al 2011 .

The DSPs belong to the protein tyrosine phosphatase (PTP) superfamily of cysteine-dependent phosphatases. Dual-specificity phosphatase domain folding is the characteristic αβα PTP fold and from *in silico* modeling it's suggested to have four or five β strands surround by α-helices[11]. This domain comprises a conserved $CysX_5Arg$

motif that hydrolyzes phosphoester bonds and cysteine functions as the nucleophile (fig. 4). In addition, DSP active site allows dephosphorylation of phosphotyrosine or phosphoserine/phosphothreonine residues. *In vivo* human laforin dephosphorylates glycogen along with the dephosphorylation of various proteins, namely GSK3β and tau[12,13,14,15].

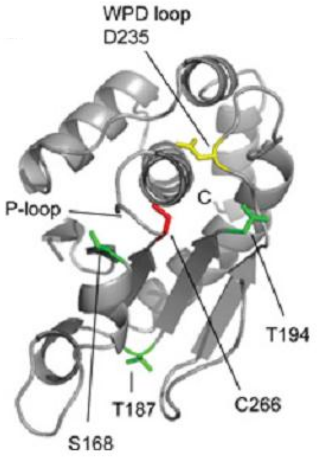

Figure 4. Model 3D structure of laforin DSP domain. The laforin DSPD (residues 157–326) was in silico modeled, using the crystal structure of human DUSP22 (PDB code: 1WRM) as a template. The position of the characteristic P-loop, containing the catalytic Cys266 residue (in red), and the WPD-loop, containing the conserved Asp235 (in yellow), is indicated. The positions of other residues described in the present study (Ser168, Thr187 and Thr194; in green) and the C-terminus are also indicated.

Regarding laforin activity, it was evidenced that dimerization through its CBM domain is important for phosphatase activity and it was observed that recombinant laforin produced in bacteria, cell culture or even tissue form dimers[16]. However the mechanism responsible for this dimerization and its real function purpose is not well understood[17,18]. Relatively to human malin, this protein is encoded by EPM2B or NHLRC1 gene of 1344 bp located in chromosome 6p22.3. The coding sequence has 1188 bp and product protein has 395 amino acids (fig. 5). Malin is highly expressed in brain, spinal cord, heart, liver, skeletal muscle and pancreas. The subcellular localization is mainly in the cytoplasm and some at the nucleus[19].

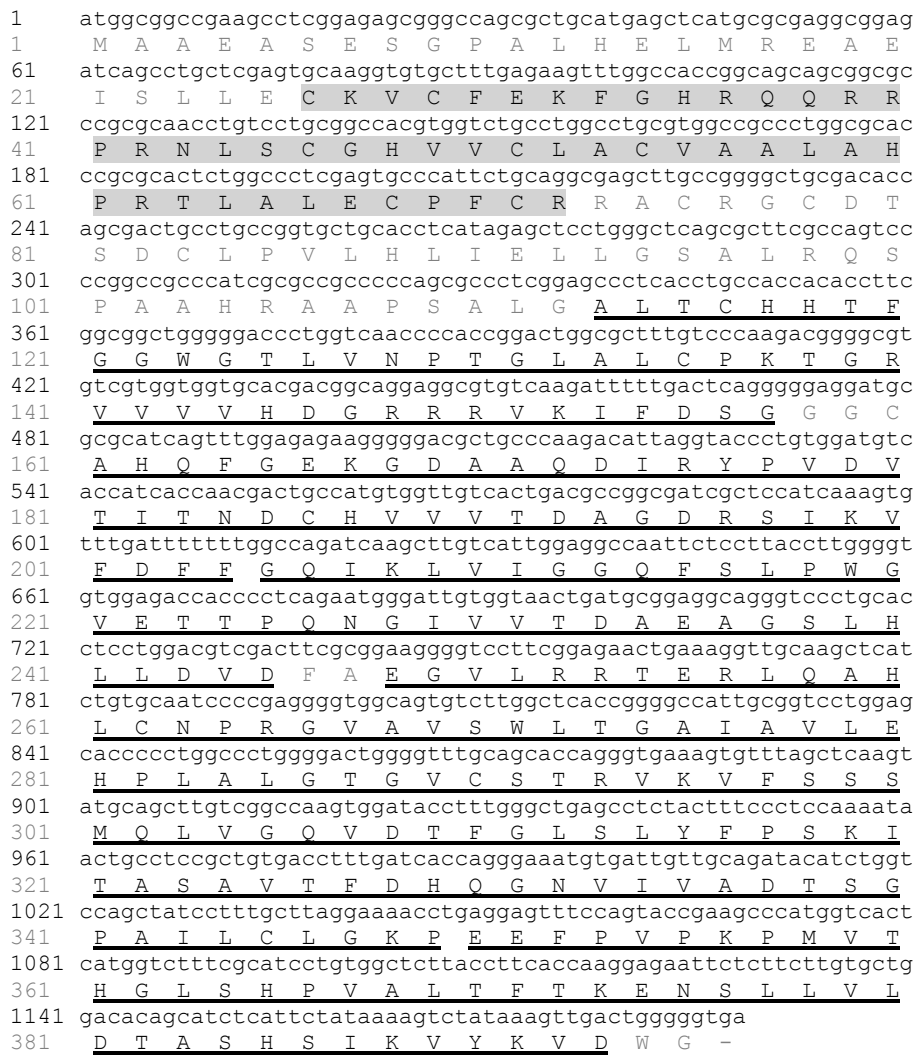

Figure 5. Coding sequence and amino acid sequence of human malin (UniProt: Q6VVB1).

Gene nucleotide sequence is denoted in black and the protein sequence is in grey. The RING finger domain is highlighted in grey and the 6 NHL motifs are underlined.

Regarding the domains, it presents a RING zinc finger motif in its N-terminal (C-X_2-C-X_{16}-C-X_1-H-X_2-C-X_2-C-X_{13}-C-X_2-C) between 26th and 72nd residues which is similar with those RING-HG types (fig. 6). That means a linear series of conserved cysteine and histidine residues, where X can be any amino acid. Two Zn^{2+} ions are coordinated to the cysteines and/or the histidines with a four coordination number. This motif folds into a compact domain with a central β-sheet and an α-helix (fig. 6)[19,20]. The RING

finger confers E3 ubiquitin ligase function to malin, being the E3 ligase role to choose specific targets for ubiquitination. This means that E3 ubiquitin ligases bind to the respective target proteins and the E2 conjugative enzymes which then promote the transfer of ubiquitin from E2 to a lysine residue of the target protein[21].

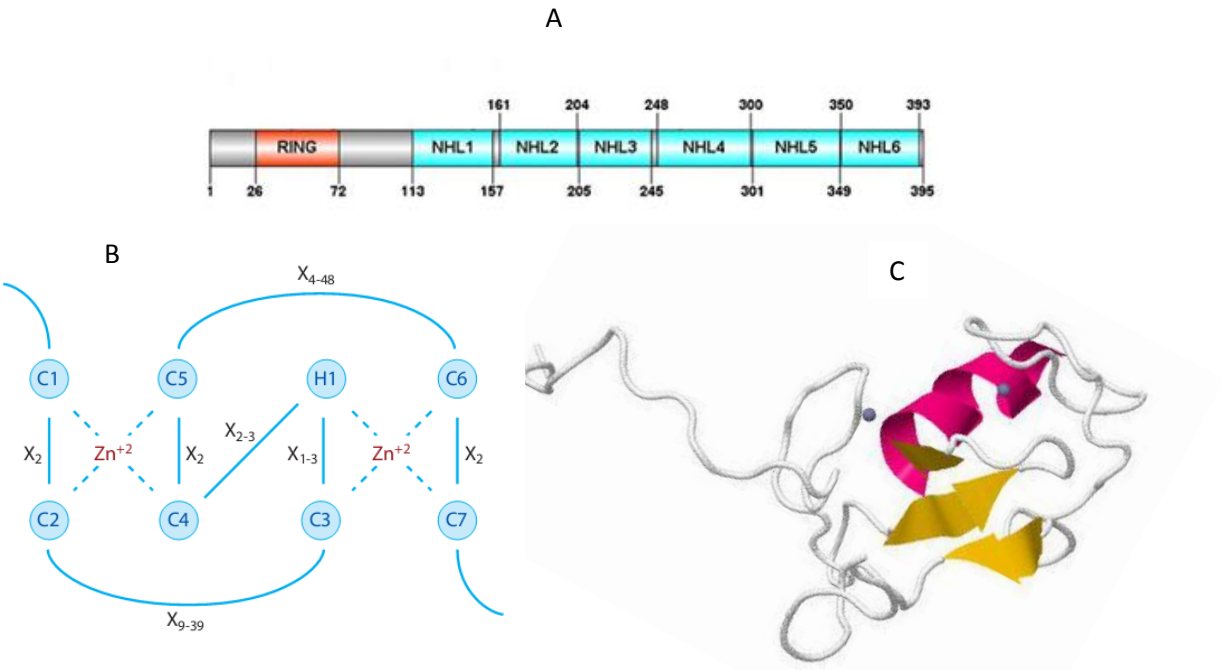

Figure 6. The Malin primary structure and models structure of RING finger |A. Schematic overview of primary structure of malin. The RING finger domain is outlined in red and the 6 NHL repeats domain ate blue. **|B**. Schematic organization of the characteristic RING-HC domain type. The first cysteine that coordinates zinc ion (II) is labeled as C1, and so on. H1 corresponds to the histidine ligand. X_n is the number of any amino acid residues in the spacer regions between the ligands. **|C.** Example of 3D model structure of the RING finger domain using tripartite motif protein 32 as template (PDB code: 2ct2A).

At the C-terminal, malin contains a NHL repeat domain, a conserved structural motif of six bladed β propeller, arranged in a radial fashion around a central axis, and each blade consists of twisted four stranded antiparallel β-sheet with an example of this domain present in Brat protein (fig. 7)[22]. This domain was first identified from three proteins, NCL-1, HT2A and LIN-41 and that's the origin of NHL name[23]. Concerning malin protein, the NHL domain is involved in protein-protein interactions[24] as it will be shown in the next topic.

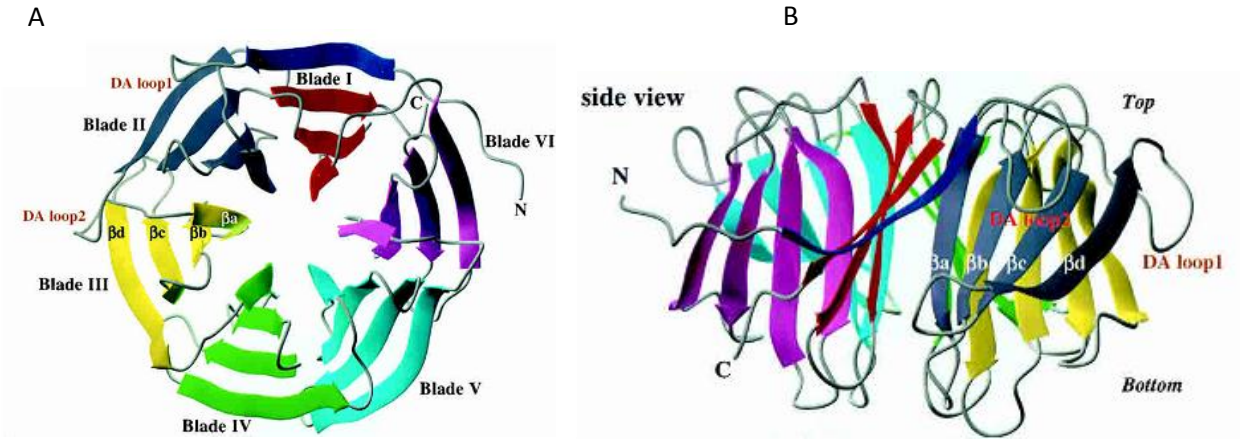

Figure 7. Brat NHL domain structure and surface. **|A** The NHL domain displayed at the "top" surface, with loops connecting the β-strands of each blade. **|B** NHL domain side view. The innermost strand of each blade is labeled 'a' and the outermost strand by 'd'. Adapted from Edwards 2003.

1.2.3.1. The role of laforin and malin in glycogen metabolism: the laforin-malin complex

Laforin functions as a mammalian glycogen phosphatase in vivo to detect, control and repair levels of glycogen phosphorylation above a certain threshold, in part due to the enzymatic error of glycogen synthase mentioned above. Otherwise, glycogen would accumulate phosphate residues and longer unit chains overtime, due to inhibited branching by the phosphates, along with the branching enzyme saturation and would eventually form insoluble polyglucosans (poorly branched glycogen particles)[25,26].

Laforin is a binding partner of malin and it was shown their interaction *in vitro* and *in vivo*. Laforin establish interactions with malin with the region between the CBM and DSP domains[27]. Both proteins form a complex, and laforin recruits substrates to be ubiquitinated by malin. These substrates are specific and many of them are enzymes involved in glycogen synthesis. Thus, besides the role as glycogen phosphatase, laforin functions also as an adaptor protein of malin targeting it to the respective substrates. The ubiquitination process consists of three steps (fig. 8). The first one is the activation of ubiquitin protein by E1 (ubiquitin activating enzyme) and transfer to E2, a conjugating enzyme. E3 ubiquitin ligase then interacts with the E2 promoting the transfer of Ub from E2 to the target protein. If a chain of at least four Ub is formed and attached to the Lys48 residue, the substrate protein is targeted for degradation in 26 S proteasome. However, mono or polyubiquitination of chains linked to the Lys63 it is a

signal of intracellular trafficking, activity modulation, cellular localization, DNA repair or signal transduction response. The outcome of ubiquitination in the cell, degradation or signaling purpose, is usually driven by ubiquitin receptors (UbR) that bind and interpret the ubiquitin signal[20].

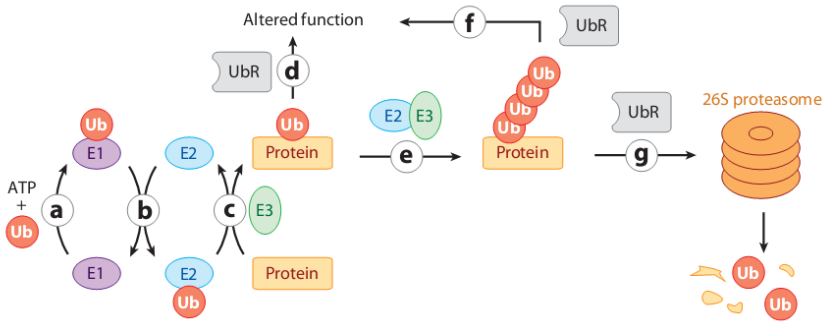

Figure 8. The ubiquitin system. a) Ubiquitin (Ub) and ubiquitin-like proteins are activated for transfer by E1(ubiquitin-activating enzyme). b) Activated ubiquitin is transferred from the active-site cysteine of E1 to the active-site cysteine of an E2 ubiquitin-conjugating enzyme. c) The E2-Ub thioester linkage interacts with an E3 ubiquitin ligase, which catalyzes transfer of Ub from E2-Ub to the substrate. Monoubiquitinated substrate can dissociate from E3 (d) or can acquire additional Ub modifications in the form of multiple single attachments (multiubiquitination, not shown) or an ubiquitin chain (polyubiquitination) (e). The chain can be attached via different lysine residues of ubiquitin. Concerning monoubiquitin and some types of chains, those assembled via Lys63 of ubiquitin serve mainly to change the protein function (f), polyubiquitin chains assembled via the Lys48 residue directs the substrate to the proteasome for degradation (g).

The laforin–malin complex binds and ubiquitinates the muscle isoform of glycogen synthase (MGS), and the protein targeting to glycogen R5/PTG[17]. These proteins are then degraded by the proteasome. There are two isoforms of glycogen synthase, one that is only expressed in liver and the muscle isoform which is expressed in muscle and the other tissues like neurons. Although neurons have the enzymes for glycogen synthesis, the process is blocked due to MGS and R5/PTG degradation ensured by the laforin-malin complex and phosphorylation[28]. In the liver, however, the laforin-malin complex only degrades the R5/PTG and not the LGS[29]. This suggests the existence of tissue-specific differences in the regulation of glycogen synthesis by the laforin–malin complex. Moreover, the complex also mediates ubiquitination of the debranching enzyme[30] and even laforin itself as control manner, that is, when laforin levels are reduced, no laforin-malin complex is formed and metabolic proteins will not be degraded (fig. 9).

Considering this, the laforin-malin complex downregulates glycogen synthesis and prevent poorly branched glycogen accumulation. Furthermore, the interaction of laforin and malin is a controlled process and enhanced through laforin Ser[25] residue desphosphorylation by AMPK. It also phosphorylates R5/PTG accelerating its ubiquitination by the complex. AMPK is a AMP dependent-kinase that functions as a metabolic sensor-protein and when activated, it upregulates catabolic pathways[3,11,29].

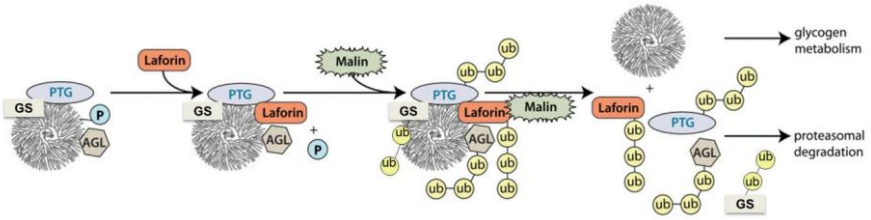

Figure 9: A model for the role of laforin-malin complex in glycogen metabolism. R5/PTG, AGL and GS bind to glycogen during its metabolism. Glycogen suffers phosphorylation (P) due to GS and laforin releases these phosphate when binds to glycogen via CBM. In addition, laforin directly interacts with PTG. Malin interacts with laforin which allows malin to ubiquitinate R5/PTG, AGL, GS and laforin signaling them to be degraded in proteasome system. Adapted from Worby et al. 2008.

Besides the role of laforin-malin complex in the regulation of glycogen metabolic enzymes, many findings support other pathways to control the activity of GS. One is through neuronatin degradation. This protein is expressed in several tissues and is involved in regulation of adipogenesis, in glucose-mediated insulin secretion at β pancreatic cells and stimulates glycogen synthase. Neuronatin inactivates GSK3β (an inhibitor of GS) through phosphorylation, which in turn allows GS to stay activated. Malin interacts with neuronatin via NHL domain promoting its ubiquitination and consequent degradation. Since degraded, neuronatin cannot inactivate GSK3β, which in turns decreases GS activity (fig. 10)[31].

Another pathway is the activation of GSK3β by direct interaction with laforin which finally suppresses GS activate state too (fig. 10)[3,11,27,29].

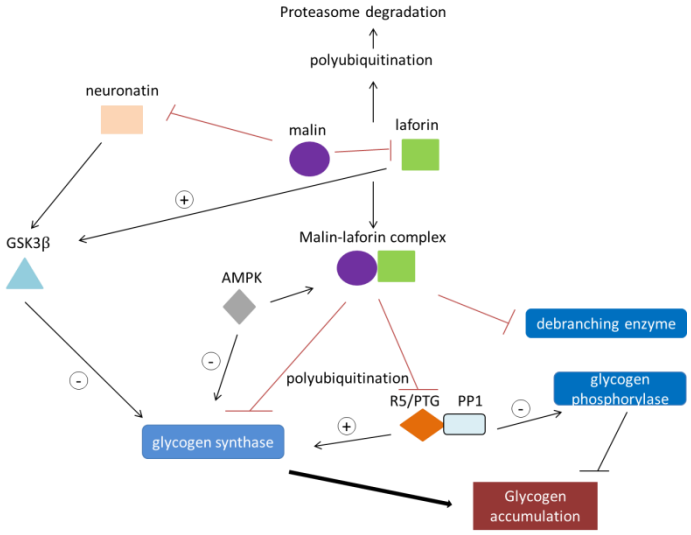

Figure 10. Representative model of different pathways of glycogen synthase downregulation by laforin and malin. Malin forms a stable complex with laforin which then polyubiquitinates debranching enzyme, glycogen synthase and R5/PTG as a signal to be degraded in proteasome. Another pathway is the ubiquitination of neuronatin that inhibits GSK3β. Thus, degraded neuronatin maintains GSK3β activated which then inhibits GS activity. Laforin can also act directly with GSK3β activating it and them GS is inhibited. Line in red refers to ubiquitination and degradation.

1.3. Lafora disease (LD):

1.3.1 Genetics

Lafora disease is an autossomal recessive neurodegenerative disorder and is the most common of the five forms of progressive myoclonic epilepsy. This disease is caused by recessive inherited mutations along the EPM2A or EPM2B genes. To date, 48 mutations were detected in EPM2A gene and 51 mutations in the EPM2B gene in LD families and has been reported that approximately 58% of LD patients have mutations in EPM2A gene and 35% present mutations in EPM2B gene. These mutations distribute more frequently in Mediterranean countries, including Spain, Italy and France, also in North Africa, Middle East and central Asia[32,33].

1.3.2 Phenotype

The first symptom usually appears during adolescence, between 12-17 years but in some cases it can be manifested in 6-year-old children[34,35]. In an early stage, the patients have generalized tonic-clonic, absence, myoclonic and visual seizures with temporary blindness and visual hallucinations. Cognitive deficits including school performance, apathy and emotional disturbance are also present and tend to be progressive. As the disease proceeds, ataxia, dysarthria, and gradually neurologic degeneration are common, paralleled by frequently myoclonic and absence seizures. Dementia sets in progressively and after 10 years from onset, patients become into a vegetative state and then die due to muscle wasting, central nervous system deterioration and status epilepticus. The estimated prevalence of Lafora disease in Europe is 1-9/10^6 because patient's life is limited[13,34,35].

Among the progressive myoclonus epilepsies, Lafora disease is clinically characterized by the presence of aggregates of glycogen-like polyglucosans called Lafora bodies. These polysaccharides are less branched and more phosphorylated, so they lack the soluble symmetry structure of the glycogen and thus they precipitate in the cytoplasm of the cell[34,35]. Lafora bodies accumulate in the central nervous system, in the perikarya and dendrites, and in another tissues such as retina, liver, cardiac and skeletal muscle, and skin[13,36].

1.3.3 Lafora disease pathogenesis: Molecular mechanisms

Laforin acts during the glycogen synthesis preventing formation of glycogen hyperphosphorylated and the development of polyglucosans formation overtime[37,38,39,40]. Indeed, it was showed that laforin is like a sensor of glycogen as it increases, the levels of laforin also increase[41]. This was confirmed from analyzes of glycogen of a mouse model of Lafora disease, Epm2a -/- mice that showed increased content of phosphate in muscle and tissue. This study was extended to demonstrate that in absence or deficiency of laforin there is a hyperphosphorylation of glycogen and tends to develop gradually into insoluble and less branched polyglucosans, which accumulate and forms the Lafora bodies typical of Lafora disease, in an age-dependent manner[26,42].

Concerning the laforin-malin complex, it was shown recently in an extended study that in Epm2b-/- mice, the levels of MGS increased in the Lafora bodies accumulated in the neurons. These mice had progressive loss of neuronal cells and propensity to suffer myoclonic seizures, which emphasizes the laforin-malin complex importance and that less branched glycogen accumulation, may also be a contributor to neurodegeneration in LD[43]. This means, the impairment of this complex (due to loss of function of one of the proteins or the interaction between them is affected) leads to less inactivation of the glycogenic targets and then to less branched glycogen accumulation. This is evident to contribute to LB formation even in neurons and probably related with neurodegeneration observed in LD, since neurons do not synthesize glycogen[28,37].

Recent studies have shown that both laforin and malin form centrosomal aggresomes (inclusions of E3 enzymes that ubiquitinate and degrade misfolded proteins when they accumulate) when proteosomal activity is inhibited, which suggests that both proteins are involved in UPS and these agressomes facilitate malin ubiquitination of it substrates to be degraded. In this way, malin-laforin complex with Hsp 70, a heat-shock protein, may contribute to rescue cells from endoplasmic reticulum stress and a deficiency in malin or laforin leads to an increase of misfolded proteins that confer toxicity to the cells, and thereby is supported that this phenomenon triggers primarily LD[44,45].

Although the carbohydrate metabolism disorder and impairment in protein clearance are likely to be the cause of LD, it was suggested that alterations in autophagy can also contribute to the pathology, since it was observed that laforin regulates autophagy process and defected laforin-malin complex in mice led to accumulation of autophagy substrates, before Lafora bodies detection. Those animals showed memory deficits and epileptiform activity similar to LD patients[46,47,48].

Moreover it was shown that laforin dephosphorylates tau protein. In the absence of laforin, in Epm2a -/- mice, tau protein became hyperphosphorylated and accumulated as NTFs (neurofibrillary tangles) which means that NTFs formation can contribute to some of the symptoms of LD like dementia, that appears in diseases resulted from abnormal regulation of tau, such as Alzheimer's disease[49].

All this findings suggests that multiple mechanisms drive the progression of LD. The disease may result from defective protein clearance and degradation, and consequently cell death, including neuron degeneration, and the appearance and

formation of Lafora bodies, however more studies are needed to understand the pathways driving those mechanisms.

1.4. Laforin and malin interaction: mapping regions involved in complex formation

Initially Gentry and co-workers mapped the domains responsible for the malin–laforin interaction, by yeast two-hybrid experiments. They have tested two constructs, malin's RING domain and NHL domain to interact with laforin and CBM or DSP domain. Only the NHL domain interacted with laforin. Later, it was shown that this region of interaction is confined to residues from 208 to 393[28].

Another study performed by Lohi had the purpose to check the critical regions involved in the interaction, with several constructs with laforin and malin domains done (fig. 11). It was concluded that the region encoded by exon 2, (102-160 residues) between the CBM and DSP domains of laforin is critical for the interaction between the two proteins.

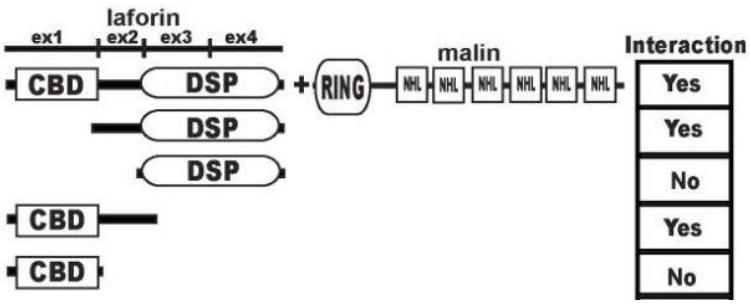

Figure 11. Representation of constructs used in mapping interaction regions in yeast two-hybrid experiments. Adapted from Lohi et al 2005.

All these interactions were determined by yeast-to-hybrid or mammalian-two-hybrid, to give the suitable conditions of post-translational changes for the proteins, and confirmed by immunoprecipitation assays.

Considering the importance of laforin and malin interactions and the stability of the respective complex, additional studies, from expression and purification to interaction characterization may be promising for structure studies, since the crystal structure of both proteins and the complex are not available yet.

2. Objectives

Laforin and malin, as a complex, play essential functions in terms of glycogen regulation and in the system responsible for the degradation of misfolded proteins. Regarding this, the main goal of the project was the optimization of the protein expression and purification conditions of malin and laforin derived constructs for later identification of a minimal amino acid sequence, within the inter-domain region of laforin, required for interaction with malin.

3. Materials and Methods

3.1. Materials

The different reagents used in the following procedures were obtained from Sigma-Aldrich.

3.2 Methods

3.2.1 PCR

The DNA coding sequence of laforin was used as template to generate and amplify eight constructs. The PCR reactions were performed using, *PfuTurbo* DNA polymerase (2,5 U), 0,5µM of each primer (forward and reverse), 25 µM dNTP, 1x enyme buffer (20 mM Tris-HCl (pH 8,8), 2 mM $MgSO_4$, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0,1 % Triton X-100, 0,1 mg/ml nuclease-free BSA). The PCR conditions were: denaturation at 95ºC during 2min, annealing at 56ºC during 30 s and extension at 72ºC, during 10 min with 30 cycles, , followed by an extra extension step at 72 ºC during 10 min.

The bands correspondent to each amplified construct were purified from agarose gel with the NZYGelpure kit (nzytech) according to manufacturer's instructions.

Table 1. Sequence of primers used to generate CBM and DSP constructs from the original full length laforin sequence.

Construct	Primers (5'→3')
CBM 120	For ATATATGCTCTTCTAGTCGCTTCCGCTTTGGGGTG Rev TATATAGCTCTTCATGCACCATCCACCAAGTTGTTTTC
CBM140	For ATATATGCTCTTCTAGTCGCTTCCGCTTTGGGGTG Rev TATATAGCTCTTCATGCCTTCATTTCATTGGTGTGCCC
CBM159	For ATATATGCTCTTCTAGTCGCTTCCGCTTTGGGGTG Rev TATATAGCTCTTCATGCTCTTGAATAATGCATGGCTTGG
DSP111	For ATATATGCTCTTCTAGTACTTACAATGAAAACAACTTGGTG Rev TATATAGCTCTTCATGCCAGGCTACACACAGAAGAAC
DSP130	For ATATATGCTCTTCTAGTATTGAGGCCACTGGGCACAC Rev TATATAGCTCTTCATGCCAGGCTACACACAGAAGAAC
DSP150	For ATATATGCTCTTCTAGTTTTAATATTGCAGGCCACCAAG Rev TATATAGCTCTTCATGCCAGGCTACACACAGAAGAAC

The recognition site of Sap I endonuclease is underlined and the cleavage site is dashed underlined.

Table 2. Sequence of the different primers used to generate CBM and DSP constructs from the codon optimized laforin coding sequence.

Construct	Primers (5'→3')
CBM 120	For ATATATGCTCTTCTAGTCGTTTTCGTTTTGGTGTTGTTG
	Rev TATATAGCTCTTCATGCACCATCCACCAGATTGTTTTC
CBM140	For ATATATGCTCTTCTAGTCGTTTTCGTTTTGGTGTTGTTG
	Rev TATATAGCTCTTCATGCTTTCATTTCATTGGTATGACCG
CBM159	For ATATATGCTCTTCTAGTCGTTTTCGTTTTGGTGTTGTTG
	Rev TATATAGCTCTTCATGCACGGCTATAATGCATTGCCTG
DSP 111	For ATATATGCTCTTCTAGTACCTATAATGAAAACAATCTG
	Rev TATATAGCTCTTCATGCCAGGCTGCAAACGCTGCTA
DSP 130	For ATATATGCTCTTCTAGTATTGAAGCAACCGGTCATACC
	Rev TATATAGCTCTTCATGCCAGGCTGCAAACGCTGCTA
DSP 150	For ATATATGCTCTTCTAGTGCAGGTCATCAGGCAATGC
	Rev TATATAGCTCTTCATGCCAGGCTGCAAACGCTGCTA

The recognition site of Sap I endonuclease is underlined and the cleavage site is dashed underlined.

3.2.2 Cloning of PCR products

The cloning of the amplified constructs was performed by fragment exchange (FX) cloning strategy[59]. This method uses a type IIS restriction enzyme, SapI, which cleaves the DNA outside the recognition site generating two overhangs with variable sequence. This allows a directional cloning of the insert by using just one endonuclease. Both PCR product and vector are digested simultaneously by the enzyme followed by ligation. An initial cloning is performed in pINITIAL vector for sequencing and then subcloning is done by digesting simultaneously the pINITIAL-derivative plus the final expression vector again with SapI. The two SapI restriction sites in the vectors flank the ccdB gene which codes a toxin functioning as a counter selection marker and is the region of the insertion of the sequence of interest. This marker plus the sacB gene, which confers sucrose sensitivity, confer a double colonies' selection to cells that just have incorporated the recombinant expression vector.

For initial cloning of PCR products, 50ng of pINITIAL vector was mixed with each amplified construct to a final molar ratio of 1:5 (vector:insert). It was added 1µL of 10x buffer 4 (New England Biolabs) and 1µL of SapI (2U) (New England Biolabs). The mixture was digested for 1h at 37ºC and the enzyme was then inactivated at 65ºC by 20 min. Ligation of the fragments was performed by 1,25µL of T4 ligase (1U/µL) (New

England Biolabs) in the presence of 0,4 µL of 25mM ATP for 1h at 25ºC followed by heat-inactivation at 65ºC ºC during 20 min.. All the mixture volume was used to transform competent TOP10F' E. coli cells. Aliquots were plated on LB-agar in the presence of 50µg/mL of kanamycin. A clone of each recombinant pINITIAL was selected for miniprep plasmid purification using the NZYminiprep (nzytech).

Each cloned laforin construct was verified by Sanger sequencing by an external sequencing service provider, either by STABVida (Oeiras – Portugal) or by Macrogen (Seoul – Korea).

3.2.3 Subcloning

The expression vectors used for subcloning of the laforin constructs are p7XNH3 and p7XC3H which are compatible with the FX cloning method. These vectors result from modified pET26a in which the two SapI sites were introduced along with the coding sequence for CcdB toxin[59]. A 10His-tag with a 3C cleavage site was added on the N-terminal giving the final p7XNH3 or at the C-terminal yielding the p7XCH vector.

After sequencing verification, each pINITIAL-laforin construct was mixed with the respective expression vector at a final molar ratio of 1:4 (expression vector:pINITIAL-derivative). In the case of laforin constructs containing the CBM domain, subcloning was performed with the p7XNH3 vector and for laforin constructs with the DSP domain it was used the p7XC3H vector. The same procedure as described in cloning section was followed, however the final recombinant expression vectors were incorporated in competent E. coli BL21 star.

3.2.4. Malin constructs

Regarding Malin, two constructs were already available in the laboratory. The first consists of malin fused with a His_6-tag and a TEV cleavage site at the N-terminal, inserted in a modified pET-28b named pSKB3. The second one, pET-GST is a modified pET21b vector, where GST coding sequence was cloned. The construct Malin_pET-GST codes for a N-terminal T7 tag, followed by GST, a thrombin cleavage site, malin and a

C-terminal His_6 tag. The next table displays these three constructs. Note all the constructs were sequenced prior to the expression experiences.

These both constructs contain a GST-tag (26 kDa) at the N-terminal of malin that besides improving malin solubility, it can also be used to purify the protein by a GST affinity chromatography. Furthermore GST-tag might also protect the protein from possible intracellular proteolysis. In addiction of GST, malin is also fused to a His-tag at the C-terminal or at the N-terminal prior of the GST. The position of His-tag at the terminal is an important factor to evaluate, because the malin native conformation may contain one or both the terminus buried inside the protein core. The construct malin in pET-GST is also fused to a T7 epitope tag at the N-terminal that was carried by the initial vector pET 21b. This T7 tag is a peptide (~4 kDa) that can be used to immunoaffinity purification, immunoblotting, immunofluorescence assays and immunoprecipitations.

Table 4 | List of the different malin fusion constructs and the scheme of the elements codified.

Construct	
Malin_pSKB3	His_6-TEV-Malin
Malin_pET-GST	T7-GST-trombin site-Malin-His_6

3.2.5 Small-scale soluble expression screenings

E. coli BL21 star was transformed with each amplified laforin derived constructs. A fresh colony from each construct was grown overnight at 37ºC in 10mL of LB medium supplemented with 50 µg/mL of kanamycin antibiotic. Each pre-inoculum was diluted 1:40 into 50mL of LB medium containing 50µg/mL of kanamycin and grown at 37ºC until $O.D_{600}$ reaches 0,5. At this point, the protein expression was induced for all the constructs with 0,1mM of IPTG and incubated overnight at 18ºC. 1mL aliquots were collected before adding IPTG (final concentration), after 3 hours of induction and after overnight induction. BugBuster Protein Extraction Reagent (Novagen) was used to extract total protein and to obtain soluble and insoluble fraction. This reagent is composed by a mixture of non-ionic detergents that enable the gentle cell wall

disruption of E. coli releasing the proteins including the target protein expressed, without denaturing the soluble protein. The presence of the endonuclease benzonase on this reagent allows the degradation of DNA and RNA giving low viscosity and clarified extracts. The protocol suggested by BugBuster manufacturer was followed. Soluble and insoluble fractions were separated by centrifugation and the insoluble fraction resuspended in 200 µL of PBS buffer 1x. Both fractions were analyzed by SDS-PAGE to evaluate the soluble form of the laforin constructs.

3.2.6 Small-scale soluble expression screenings in the presence of oxidases and foldases

Malin_PSK3 and Malin_pET-GST constructs were screened by expression in E. coli BL21 star which co-expresses an oxidase and a foldase[57]. The transformed cells were grown overnight at 37ºC in 10mL of LB medium with 34 µg/mL of chloramphenicol to select the cells with the expression vector that codes the oxidase and foldase and 50 µg/mL of ampicillin or 50 µg/mL of kanamycin concerning the malin construct. The pre-inoculum was diluted 1:40 into 50mL of LB medium containing 34 µg/mL of chloramphenicol and 50 µg/mL of ampicillin or 50µg/mL of kanamycin and grown at 37ºC until reaches an $O.D_{600}$ of 0,5. Then 0,5% of L-arabinose was added 30 minutes before the addiction of 0,1mM IPTG, to pre-induce the expression of the foldases and oxidase. The cell culture was then incubated overnight at 18ºC. BugBuster reagent was used to extract all the protein and to obtain the soluble and insoluble fractions as mentioned above. The collected fractions were analyzed by SDS-PAGE and by Western blotting to determine the best expression condition with higher malin soluble form.

3.2.7 Large-scale expression of recombinant protein

A pre –inoculum of transformed E. coli BL21 star with each cloned laforin construct and with malin constructs were grown overnight at 37ºC in LB medium supplemented with 50 µg/mL of kanamycin or 50 µg/mL of ampicillin . LB medium was inoculated with 1:40 of the pre-inoculum and grown at 37ºC until reaching an OD_{600} of 0,5. At this point, expression was induced by addition of 0,1mM of IPTG and carried out overnight

at 18°C. The induced cells were then harvested by centrifugation (Beckman Avanti J-26 XPI, JLA 8.1000 rotor, 6000 rpm for 30 minutes at 4°C), resuspended in 10 mL per 1 L of culture with binding buffer (20 mM sodium phosphate buffer pH 7,5 with 20 mM imidazole and 0,5 M NaCl for His-tagged protein or PBS 1x pH 7,4 containing 1 mM phenylmethylsulfonyl fluoride, 2 mM EDTA and 1 mM DTT for GST-tagged protein). 600µL of DNAse I and 1mL 1M $MgCl_2$ were added to the suspension and cells were disrupted mechanically by an Emulsiflex C3 homogeneizer (Avestin). Following that, the lysed cell cultures were centrifuged (Beckman Avanti J-26 XPI, JA 25.50 rotor, 15 000rpm for 30 minutes at 4°C) to remove cell debris and inclusion bodies. The supernatant was filtered through 0,2 µm filter and applied to an affinity chromatography column.

3.2.8 Large-scale co-expression of recombinant protein with pre-expression of oxidases and foldases

E. coli BL21 cells which pre-express the oxidase and foldase and transformed with malin_pSKB3 and malin_pET-GST constructs were grown at 37ºC in LB medium in the presence of 34 µg/mL of chloramphenicol and 50 µg/mL of ampicillin or 50µg/mL of kanamycin respectively. LB medium was then inoculated with 1:40 of the pre-inoculum and grown at 37ºC until reaching an OD_{600} of 0,5. The pre-induction of the oxidase and foldases was performed by adding 0,5% L-arabinose. 30 minutes later, protein expression was induced by 0,1mM of IPTG and cell cultures were incubated overnight at 18ºC.

The induced cells were then harvested by centrifugation (Beckman Avanti J-26 XPI, JLA 8.1000 rotor, 6000 rpm for 30 minutes at 4°C), resuspended in 10 mL per 1 L of culture with binding buffer (20 mM sodium phosphate buffer pH 7,5 with 20 mM imidazole and 0,5 M NaCl for His-tagged protein or PBS 1x pH 7,4 containing 1 mM phenylmethylsulfonyl fluoride, 2 mM EDTA and 1 mM DTT for GST-tagged protein). DNAse I and $MgCl_2$ were added to the suspension and cells were disrupted mechanically by an Emulsiflex homogeneizer. Following that, the lysed cell cultures were centrifuged (Beckman Avanti J-26 XPI, JA 25.50 rotor, 15 000rpm for 30 minutes

at 4°C) to remove cell debris and inclusion bodies. The supernatant was filtered through 0,2 µm filter and immediately applied to an affinity chromatography column.

3.2.9 Purification of His-tagged recombinant protein

The presence of imidazole is important during cell lysis to reduce nonspecific interactions of host cell proteins to the column when protein is loaded into the column.

The recombinant protein in the supernatant was purified by immobilized metal ion affinity chromatography, using nickel HisTrap™ HP 5 mL column (Amersham Biosciences). The column was previously equilibrated in the binding buffer, using a peristaltic pump (Bio-Rad) at a flow rate of 5 mL/min The clarified cell extract was loaded into the column at a flow rate of 0,5 mL/min and the column was washed at a flow rate of 5 mL/min with binding buffer in a low-pressure system chromatography (Bio-Rad BioLogic LP) until A_{280} reached the baseline. Elution was performed at 5 mL/min with four elution steps with increasing concentrations of imidazole (50 mM, 100 mM, 300 mM and 500 mM).

3.2.10 Purification of GST-tagged recombinant protein

The supernatant with the recombinant protein is previously treated with Triton X-100 1% and then applied to a 5 mL GSTrap™ HP column equilibrated in PBS pH 7,4 buffer (140 mM NaCl, 2,7 mM KCl, 10 mM Na_2HPO_4, and 1,8 mM KH_2PO_4) at a flow rate of 0,5 mL/min. Wash step was performed at a flow rate of 5 mL/min with the binding buffer in a low-pressure system chromatography (Bio-Rad BioLogic LP) until A_{280} reached the baseline. Elution was carried out in a single step using 50 mM Tris-HCl, 10 mM reduced glutathione, pH 8,0 as elution buffer.

3.2.11 Purification by Size exclusion chromatography

The selected fractions eluted from the HisTrap™ HP column were pooled and then concentrated in an Amicon Ultra Centrifugal Filter Unit concentrator (Millipore) (molecular cutoff of 3 kDa) to a final volume of 700µL. The concentrated pool was then applied to a Superdex 200 10/300 GL column (GE Healthcare) previously equilibrated in 20mM sodium phosphate pH 7,5 with 150 mM NaCl, and the size-exclusion chromatography was performed at a flow rate of 0,4 mL/min, in a fast-protein liquid chromatography (FPLC) system (Bio-Rad BioLogic DuoFlow), collecting fractions of 1mL.

3.2.12 Recombinant protein analysis by Analytical size exclusion chromatography

Analytical size exclusion chromatography was performed in high-pressure liquid chromatography (HPLC) Shimadzu system equipped with an automated sample injection and a photo diode array detector, using a Superdex 200 5/150 GL analytical column previously equilibrated in 20 mM Tris pH 8,0 and 150 mM NaCl). Column calibration was performed by the manufacturer's instructions using gel filtration calibration kits (GE Healthcare). 50 µL of dialyzed and concentrated protein sample was applied to the column. Separation was performed at a flow rate of 0,4 mL/min.

3.2.13 Total protein quantification

Total protein quantifications were performed using the Thermo Scientific Pierce® Micro-BCA Protein Assay Kit (Thermo Scientific), accordingly to the standard protocol described in the instructions manual. The plates were read in a microplate UV-Vis reader (PowerWave XS Microplate Spectrophotometer, Biotek®).

3.2.14 SDS-PAGE and Western blotting

SDS-PAGE was performed in 12,5% polyacrylamide gels. All samples were 1:1 diluted in 2x loading solution (250 mM Tris-HCl pH 7,4, 8 M Urea, 4 % SDS, 1,76 % β-

mercaptoethanol and 0,02% Bromophenol Blue) followed by 10 minutes denaturation at 90 °C. The gels were run in a MiniProtean 3 system (Bio-Rad) at room temperature using a constant voltage of 150 V. The running buffer was 100 mM Tris, 100 mM Bicine with 0,1% SDS. The gel was stained by 0,2% Coomassie Brilliant Blue R in 10 % acetic acid and 50 % methanol and destained with 5% acetic acid, 25% methanol solution until background staining was cleared.

For Western blotting analysis, polyacrylamide gels were electrotransferred onto PVDF membranes (Roche) previously activated in methanol and equilibrated in transfer buffer (25 mM Tris, 192 mM Glycine, 20 % methanol). The protein transference was performed overnight at 4°C using a constant voltage of 40 V in transfer buffer, using a Trans-Blot® Electrophoretic Transfer cell apparatus (Bio-Rad). PDVF membranes were blocked for one hour with TBS-T buffer (150 mM NaCl, 10 mM Tris, pH 8,0, 0,1% Tween 20) containing 5% of skim milk. Membranes were then incubated with His tag primary antibody (mAB, Mouse – GenScript) with a 1:10 000 dilution in TBS-T buffer containing 0,5% of skim milk for one hour at room temperature. Membrane was then washed at least 5 × 5 minutes in TBS-T buffer containing 0,5% of skim milk. The incubations with the secondary antibody were also performed with TBS-T buffer with 0,5% skim milk. The secondary antibody used in these incubations was anti-mouse IgG + IgM antibody alkaline phosphatase linked (GE Healthcare) in a 1:10 000 dilution. After a washing process of at least 5 x 5 minutes in TBS-T buffer, membranes were incubated with ECF™ substrate for a maximum of 5 minutes and finally revealed by fluorescence detection in a Molecular Imager FX (Bio-Rad).

3.2.15 Sequence analysis

Protein secondary structure prediction of Laforin by three different software: PSIPRED60,61, SABLE62,63,64,65 and Jpred366.

PSIPRED server link: http://bioinf.cs.ucl.ac.uk/psipred/

SABLE server link: http://sable.cchmc.org/

Jpred3 server link: http://www.compbio.dundee.ac.uk/www-jpred/

4. Results and Discussion

4.1. Laforin constructs cloning

To achieve the objective of identifying the minimal laforin amino acid region responsible for malin interaction, a series of 6 laforin constructs covering the inter CBM-DSP domain were obtained for later interaction studies. The combination of the interactions between malin and the protein domains coded by these six constructs (represented in fig. 12) would allow the identification of the minimal 10-amino acid blocks required for such interaction.

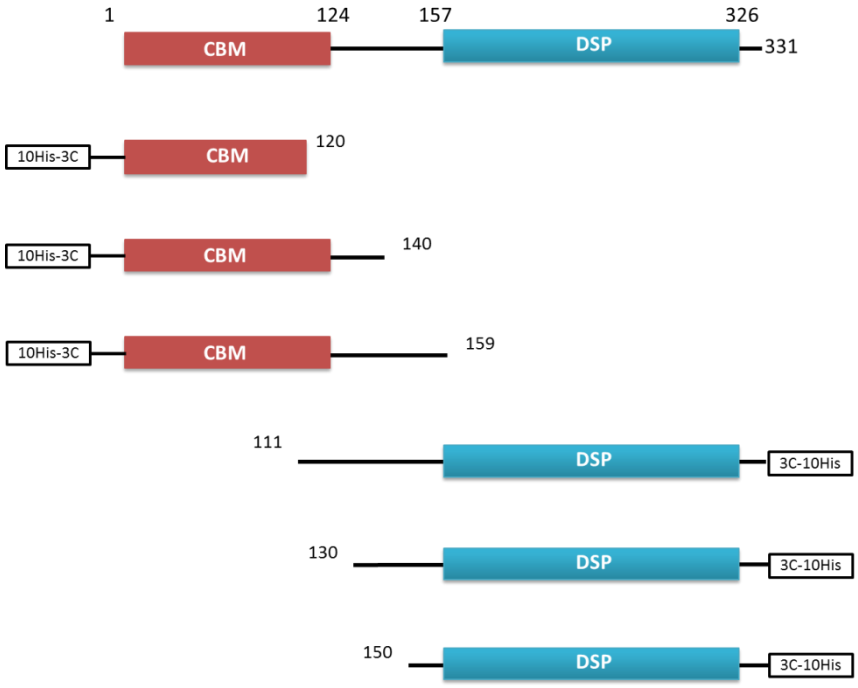

Figure 12. Schematic view of the final six constructs. The first sequence corresponds to full length laforin**.** The next three DNA constructs coding for the CBM domain, with a His_{10}-tag and 3C cleavage site at the N-terminal. The last three constructs code for the DSP domain, fused with a His_{10}-tag and a 3C cleavage at the C-terminal. The restriction sites of SapI endonuclease are displayed by arrows, used for cloning and subcloning steps.

Three constructs contain the CBM domain extended with 20 and 40 amino acids and fused to a His_{10}-tag and a 3C cleavage site. The other three constructs contain the DSP domain along with 20 and 40 amino acids, a His_{10}-tag and a 3C cleavage site fused at the C-terminal. The CBM construct which extends to the 159 residue and the DSP construct that lacks the first 111 residues would function as positive controls for interaction analysis. The presence of the His-tag facilitates the purification procedure and allows the identification of each laforin construct by using an antibody against His-tag. Moreover, this fusion strategy was aimed to prevent the possible interference of the tag with the interaction region of these constructs with malin.

The traditional cloning strategies involving two restriction enzymes and ligation step are somehow limited. The sequential steps of amplified DNA product and vector digestion, ligation and consequent transformation with intermediate purifications of the products are labor intensive and time consuming. In addition, when cloning method is set up for several different DNA sequences, the traditional cloning is limited by the possible occurrence of restrictions sites in those sequences. Therefore the FX-cloning strategy developed by Eric Geertsma and Raimund Dutzler was chosen due to its advantages (fig 13). This method is based on the use of type IIS restriction endonucleases, in particular SapI, which recognizes a non-palindromic sequence and cleaves at a fixed distance outside the recognition site.

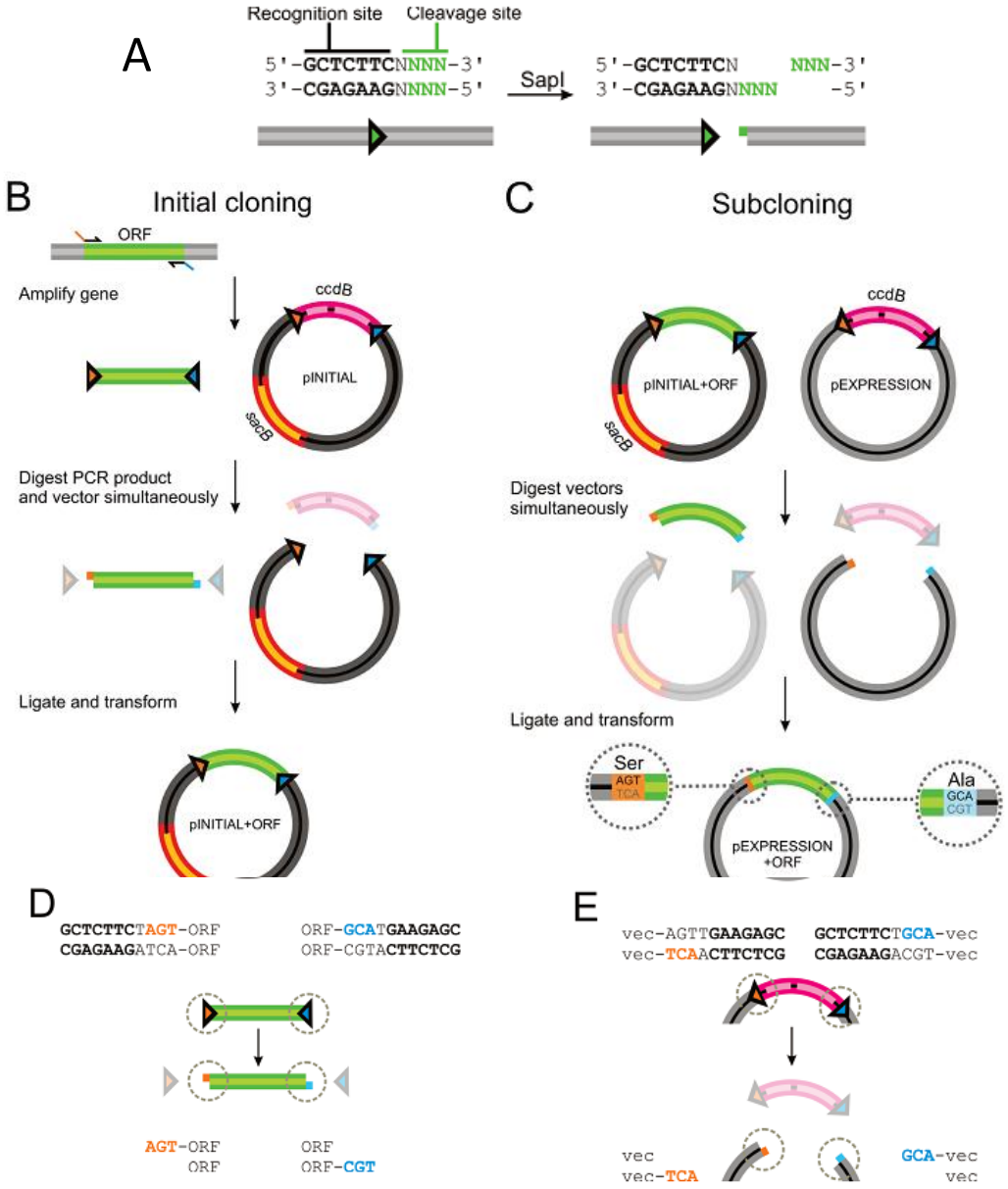

Figure 13. Schematic representation of FX cloning strategy. |A. SapI restriction site. Letters in bold correspond to the recognition site and letters in green the cleavage site. **|B.** Cloning of the PCR product to the pINITIAL vector. The amplified product displayed in green with the overhangs colored. The genes coding for the counterselection markers ccdB and sacB at the pINITIAL are outlined in pink and orange respectively. **|C.** Subcloning of the insert into an expression vector. The three nucleotides added to the sequence terminus are outlined at circles. **|D.** Orientation of the Sap I cleavage sites in the amplified product and pINITIAL vector and in the pEXPRESSION vector **(E)**. Adapted from Eric R. Geertsma et al. 2011.

The choice of SapI was based on commercial availability, with long and low frequency occurring recognition sequences and producing three base-pair overhangs. Thus, SapI originates 5' cohesive overhangs of three nucleotides of variable sequence, which allows directional cloning and prevents self-ligation of the vector using only one restriction enzyme. Both PCR product and vector are digested simultaneously by the enzyme followed by ligation. An initial cloning is performed in pINITIAL vector (fig. 13B) for sequencing as mentioned above and then subcloning is done by digesting simultaneously the pINITIAL-derivative plus the final expression vector with SapI (fig. 13C).

Moreover these triplets, in this case, AGT and GCA were chosen, code small and uncharged amino acids, a serine and alanine respectively and thus, the final protein will only have one additional amino acid at each terminal that should not interfere with protein structure. The presence of toxin ccdB in both original pINITIAL and expression vectors, prevents the appearance of false positive clones in each cloning step, resulting from self religated vector or incomplete vector digestion, keeping this strategy fast, uniform, robust and efficient.

The laforin constructs were generated by PCR using full-length laforin coding sequence as template (fig. 2). The PCR primers (table 1) used to amplify the constructs introduced different SapI sites allowing for the PCR products to be directionally cloned into the pINITIAL and subcloned into the expression vectors. An initial PCR reaction was tested using *Pfu Turbo* DNA polymerase (Stratagene). The results showed no amplification or very low yields of amplification for the different constructs. Since this, several PCR optimizations were performed to circumvent this problem. It was tested the amplification by two different DNA polimerases: the *Pfu Turbo* DNA polymerase and Taq polymerase. The presence of DMSO, an organic additive which stabilizes the DNA template and primers minimizing formation of secondary structures[50,51] was tested and touchdown PCR approach was performed to avoid the amplification of nonspecific sequences, enhancing the specificity of the initial primer–template duplex formation and thus the specificity of the final PCR product[52]. After successive attempts

by changing PCR conditions, good amplification yields were obtained. In fig. 14, is displayed an example of this result.

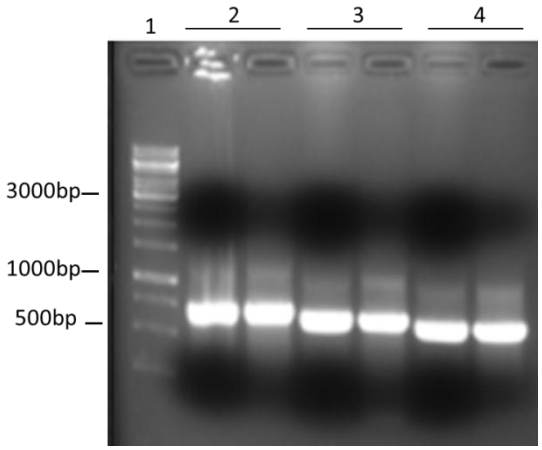

Figure 14. Agarose gel electrophoresis of DSP 111, 130 and 150 constructs amplification. Lane 1 corresponds to 1 kb DNA Ladder. Lanes 2, 3 and 4 correspond to duplicates of DSP 111, DSP 130 and DSP 150 constructs respectively. 20µL of each PCR product were applied in each lane.

The band correspondent to each amplified construct was excised from the agarose gel and DNA was purified. The amplified laforin constructs were then cloned into the pINITIAL vector for sequencing analysis following the FX cloning strategy. The respective sequencing results of most of the laforin constructs failed consecutively as shown in figure 15.

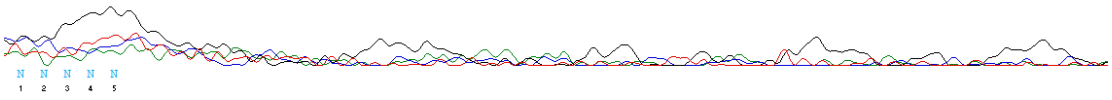

Figure 15. Failed Sequencing electrophorectogram of part of the CBM 120 construct cloned into pINITIAL vector. This is an illustrative result of the failed DNA sequencing obtained with wild-type laforin sequence constructs (Visualized on ContigExpress – Vector NTI Advance 11.0).

Successive attempts of cloning were performed to overcome this limitation, however the sequencing experiments were consistently unsuccessful. The possible reason for this problem is on the high GC content present along the first half of laforin DNA sequence (fig. 2). It is known that GC rich regions (greater than 60%) of the template DNA are difficult to amplify and to sequence under standard conditions. To overcome this problem, a codon and GC content-optimized laforin coding sequence was obtained from Invitrogen using the Geneart Gene Optimizer® Process with multi-parameter gene optimization. The optimized laforin coding sequence is shown in fig. 16 and was used as template for the following PCR amplifications. Primers are listed in table 2.

```
1    atgcgttttcgttttggtgttgttgttccgcctgcagttgccggtgcacgtccggaactg
1     M  R  F  R  F  G  V  V  V  P  P  A  V  A  G  A  R  P  E  L
61   ctggttgttggtagccgtcctgaactgggtcgttgggaaccgcgtggtgcagttcgtctg
21    L  V  V  G  S  R  P  E  L  G  R  W  E  P  R  G  A  V  R  L
121  cgtccggcaggtacagcagccggtgatggtgcactggcactgcaagaaccgggtctgtgg
41    R  P  A  G  T  A  A  G  D  G  A  L  A  L  Q  E  P  G  L  W
181  ctgggtgaagttgaactggcagccgaagaggcagcacaggatggtgcagaacctggtcgt
61    L  G  E  V  E  L  A  A  E  E  A  A  Q  D  G  A  E  P  G  R
241  gttgatacctttggtataaattcctgaaacgtgaaccgggtggtgaactgagctgggaa
81    V  D  T  F  W  Y  K  F  L  K  R  E  P  G  G  E  L  S  W  E
301  ggtaatggtccgcatcatgatcgttgttgtacctataatgaaaacaatctggtggatggt
101   G  N  G  P  H  H  D  R  C  C  T  Y  N  E  N  N  L  V  D  G
361  gtttattgtctgccgattggtcattggattgaagcaaccggtcataccaatgaaatgaaa
121   V  Y  C  L  P  I  G  H  W  I  E  A  T  G  H  T  N  E  M  K
421  cataccaccgacttctacttcaatattgcaggtcatcaggcaatgcattatagccgtatt
141   H  T  T  D  F  Y  F  N  I  A  G  H  Q  A  M  H  Y  S  R  I
481  ctgccgaatatttggctgggtagctgtccgcgtcaggttgaacatgttaccattaaactg
161   L  P  N  I  W  L  G  S  C  P  R  Q  V  E  H  V  T  I  K  L
541  aaacatgaactgggcattaccgcagtgatgaattttcagaccgaatgggatattgttcag
181   K  H  E  L  G  I  T  A  V  M  N  F  Q  T  E  W  D  I  V  Q
601  aatagcagcggttgtaatcgttatccggaaccgatgacaccggataccatgattaaactg
201   N  S  S  G  C  N  R  Y  P  E  P  M  T  P  D  T  M  I  K  L
661  tatcgtgaagaaggcctggcctatatttggatgccgactccggatatgagcaccgaaggt
221   Y  R  E  E  G  L  A  Y  I  W  M  P  T  P  D  M  S  T  E  G
721  cgtgttcagatgctgccgcaggcagtttgtctgctgcatgcactgctggaaaaaggtcat
241   R  V  Q  M  L  P  Q  A  V  C  L  L  H  A  L  L  E  K  G  H
781  attgtttatgtgcattgtaatgccggtgttggtcgtagcaccgcagccgtttgtggttgg
261   I  V  Y  V  H  C  N  A  G  V  G  R  S  T  A  A  V  C  G  W
841  ctgcagtatgttatgggttggaatctgcgtaaagtgcagtattttctgatggcaaaacgt
281   L  Q  Y  V  M  G  W  N  L  R  K  V  Q  Y  F  L  M  A  K  R
901  cctgccgtgtatattgatgaagaagcgctggcacgcgcacaagaagatttttttcagaaa
301   P  A  V  Y  I  D  E  E  A  L  A  R  A  Q  E  D  F  F  Q  K
961  tttggtaaagtgcgtagcagcgtttgcagcctgtaa
321   F  G  K  V  R  S  S  V  C  S  L  -
```

Figure 16. Sequence of the codon-optimized laforin gene, coding for the 311 amino acid sequence. The underlined sequence corresponds to the CBM domain and the dashed sequence is the DSP domain.

With the codon optimized laforin sequence the problem was circumvented, since high yields of amplification (fig 17) were obtained for each laforin construct along with readily sequencing results (fig. 18).

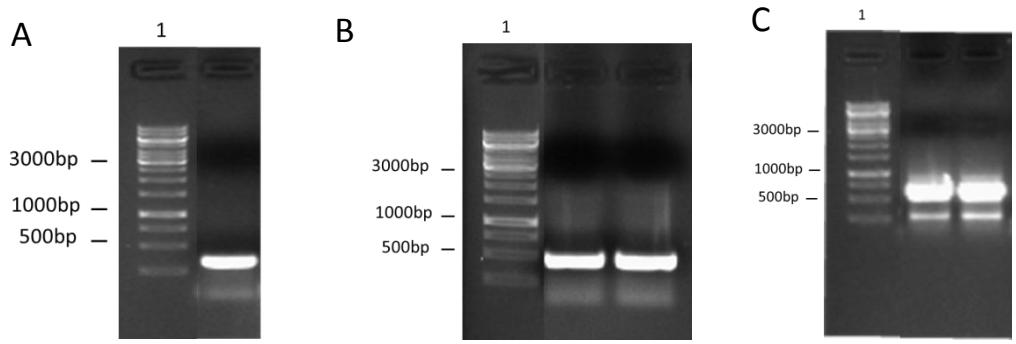

Figure 17. Agarose gel electrophoresis of CBM 120, CBM 140 and DSP 130 amplified constructs. Lane 1 corresponds to 1 kb DNA Ladder. A. Lane 2 corresponds to CBM 120 construct. B. Lane 2 and 3 corresponds to duplicates of CBM 140. C. Lanes 2 and 3 corresponds to duplicates of DSP 130. 20µL of each PCR product were applied in each lane.

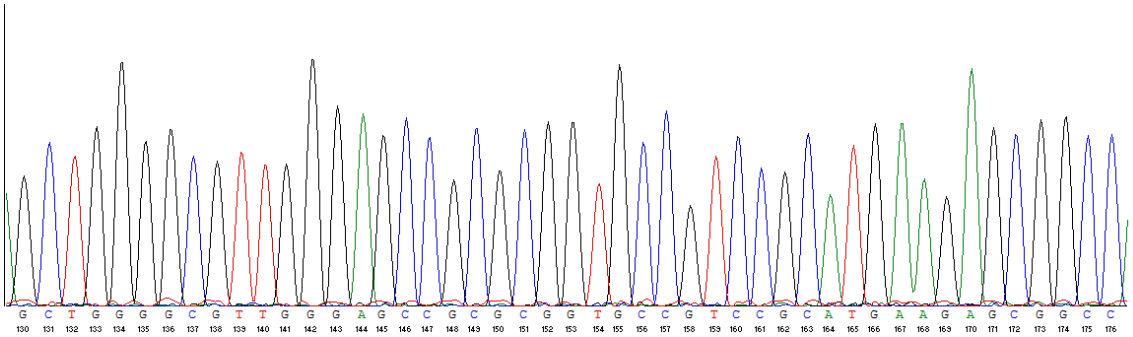

Figure 18. Part of sequencing electrophorectogram of CBM 140 in pINITIAL. The letters correspond to the initials of the nucleotides. Below is represented part of the sequence in base pairs. (Visualized on ContigExpress – Vector NTI Advance 11.0).

Indeed, when comparing the GC content of the original laforin sequence with the optimized sequence, as displayed in the fig. 19, it is observed that the first half of the wild-type sequence has high GC content with percentages reaching around 90%. These values were markedly decreased through optimization.

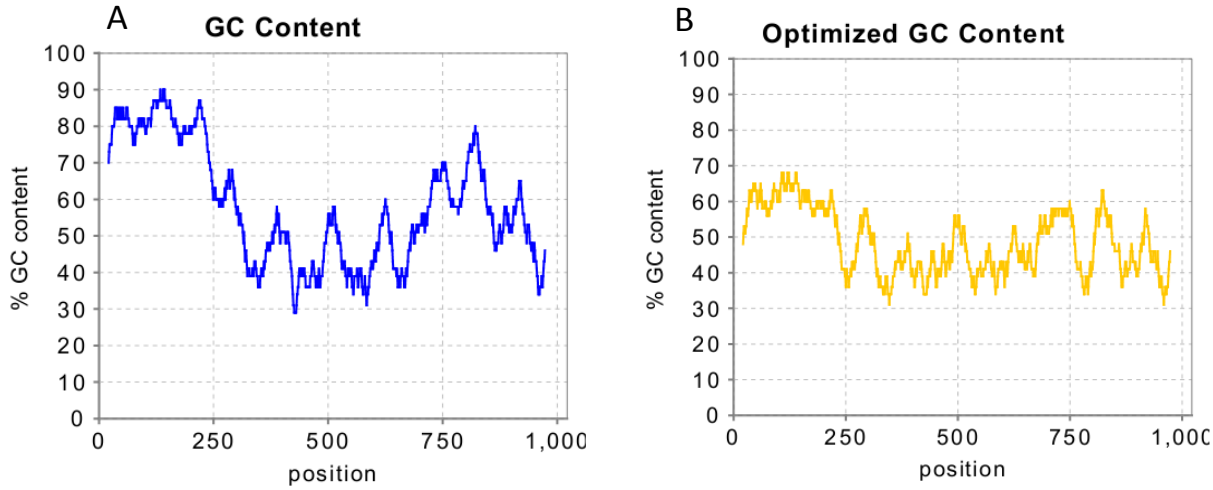

Figure 19. Plots showing the GC content of the original laforin sequence (A) and the optimized laforin sequence (B). The X axis represents the nucleotide position.

The next plots demonstrate the significant increase in codon quality of the optimized laforin sequence from the original one which may be favorable for further successful laforin constructs expression.

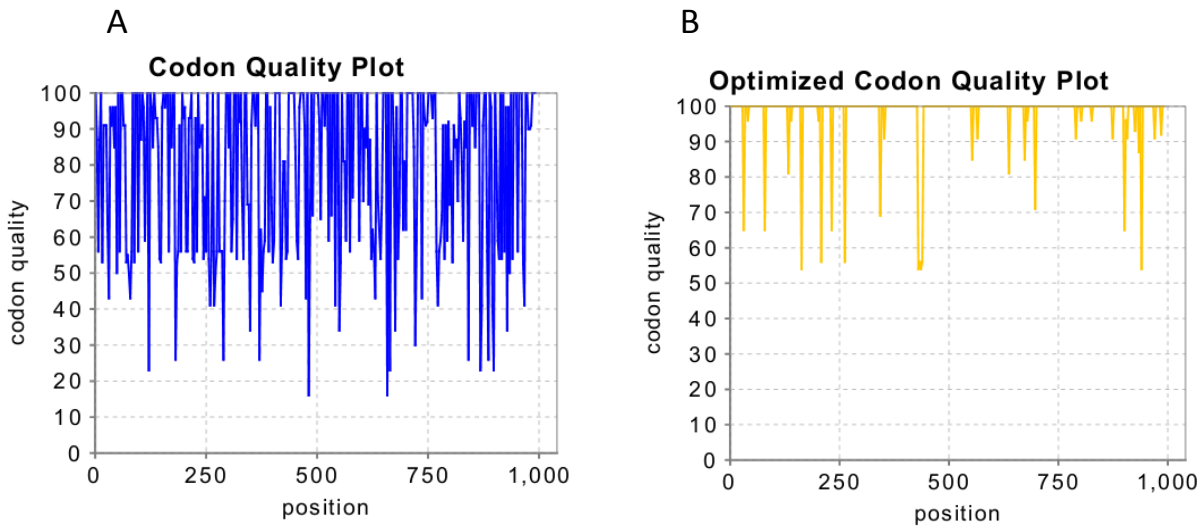

Figure 20. Plots showing the quality of used codon of the original laforin sequence (A) and the optimized laforin sequence (B). The X axis represents the nucleotide position.

After DNA sequencing analysis, laforin constructs were subcloned into the expression vectors p7XNH3 and p7XC3H. These vectors were adapted from the FX cloning strategy and were developed based on pET28a backbone, coding respectively for an N-terminal decahistidine tag followed by a 3C recognition site and the gene of interest, or for the gene of interest followed by a 3C recognition site followed by a decahistidine tag.

4.2. Small-scale expression screening of laforin constructs

Constructs coding for 120, 140 and 159 and coding for DSP 111 and 130 were transformed into BL21 star cells and expression trials were performed with expression volumes of 50 mL of LB/kan, carried out at 18 ºC with expression being induced by 0,1 mM IPTG. At this temperature the transcription and translation rates are slower, which may avoid insoluble protein aggregation. Moreover, the p7X series vectors contains the T7lac promoter and lacI gene which codes the lac repressor that suppresses the basal transcription of T7 RNA polymerase in the host genome and at the level of the vector itself, inhibiting the expression of the target gene by T7 RNA polymerase. Only after IPTG addiction, the bacteria start the production of the T7 RNA polymerase and consequent target protein expression. After expression, cells were harvested and disrupted by BugBuster^{TM}Protein Extraction Reagent (Novagen). This reagent is a mixture of non-ionic detergents which allows the gentle chemical disruption of the cell wall, without denaturating the protein. Then, soluble and insoluble fractions were obtained after centrifugation.

Both soluble and insoluble fractions collected at the end of expression, along with total protein samples collected before IPTG induction and after protein expression were run on SDS-PAGE gel (fig. 21). This allows not only to address if protein expression is efficiently repressed until IPTG induction but also to address if protein was expressed in soluble form or accumulated as inclusion bodies.

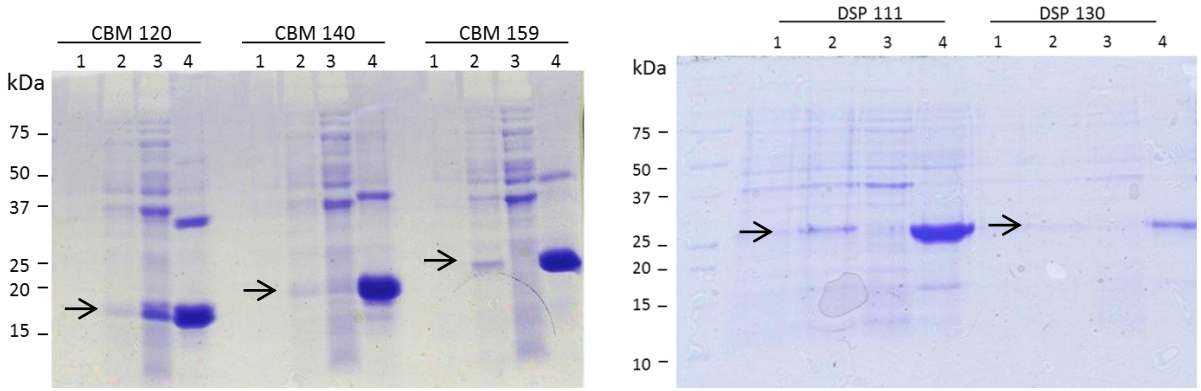

Figure 21. SDS-PAGE analysis of CBM and DSP constructs small scale expression screening. Overnight expression of different constructs in BL21 star cells, grown in LB medium and protein expression induced by 0,1mM IPTG at 18ºC. 20µL of sample was applied in each lane. Gel was stained with Comassie Blue. **1**-Total protein prior to IPTG induction; **2**-Total protein after 3h of IPTG induction; **3**-Overnight expression soluble fraction and **4**- Overnight expression insoluble fraction. The arrows correspond to the expected molecular weight of each protein. CBM 120-15,5 kDa, CBM 140- 17,5 kDa, CBM 159- 20 kDa, DSP 111- 25,3 kDa, DSP 130- 23 kDa.

As it is observed in fig. 21, all constructs were expressed and highly accumulated in inclusion bodies (Lanes 4), only CBM 120 shows significant expression in the soluble form (CBM 120 – Lane 3). It is clear that the presence of additional amino acids from both CBM and DSP domains contributes to the improper folding of those proteins leading to consequent accumulation as inclusion bodies. One possible reason is due to the type of secondary structure of this region. Secondary structure predictions of the inter-domain region shows a random coil type conformation that is expected to be exposed and may difficult the proper protein folding, leading to aggregation. In addition, it is known that small variations in few amino acid residues at the boundaries of protein structural domains can dramatically affect the yield, solubility and stability of the final constructs[53]. Taking these results into account and CBM 120 was obtained in soluble form, scale-up expression was carried out for this protein.

4.3. Large-scale soluble expression and purification of CBM 120 protein

For the scale-up expression of CBM 120 the same conditions of the screening were used as an attempt to obtain soluble expression at large quantities, without changing parameters such as E. coli strain, media, IPTG concentrations temperature and time of expression. For each construct, E. coli BL21 star strain was transformed with the expression construct and protein expression was performed in 4L of LB medium. Expression was induced with 0,1 mM IPTG (final concentration) and cells were cultured overnight at 18˚C.

Protein purification was performed by immobilized metal affinity chromatography (IMAC), which is based upon the affinity of the histidine tags for metal divalent cations, such as Ni^{2+}, Co^{2+} or Cu^{2+}. Both proteins were eluted from the column by a four – step imidazole elution (50, 100, 300 and 500mM). The chromatogram obtained by the IMAC purification is shown in fig. 22. Fractions representative of each elution step were analyzed by SDS-PAGE and western blot (fig. 23).

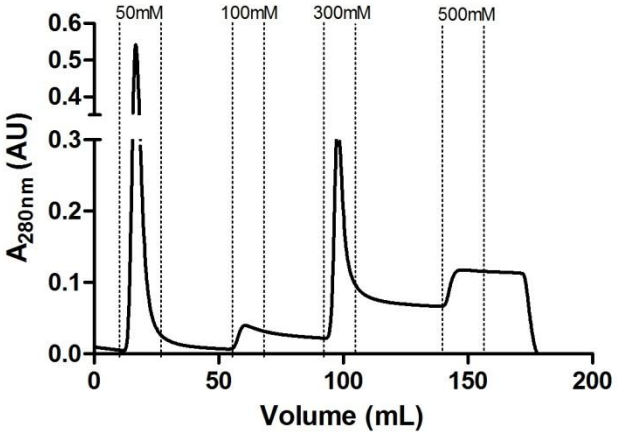

Figure 22. Chromatogram of purification of CBM 120 by immobilized metal affinity chromatography. Soluble fraction from 4L expression were loaded into a HisTrap HP 5mL column previously equilibrated in 20mM imidazole, 500mM NaCl and 20mM sodium phosphate buffer at pH 7.5. Protein elution was carried out in a four-step imidazole gradient - 50, 100, 300 and 500mM. Vertical lines outline the collected fractions in each imidazole elution step.

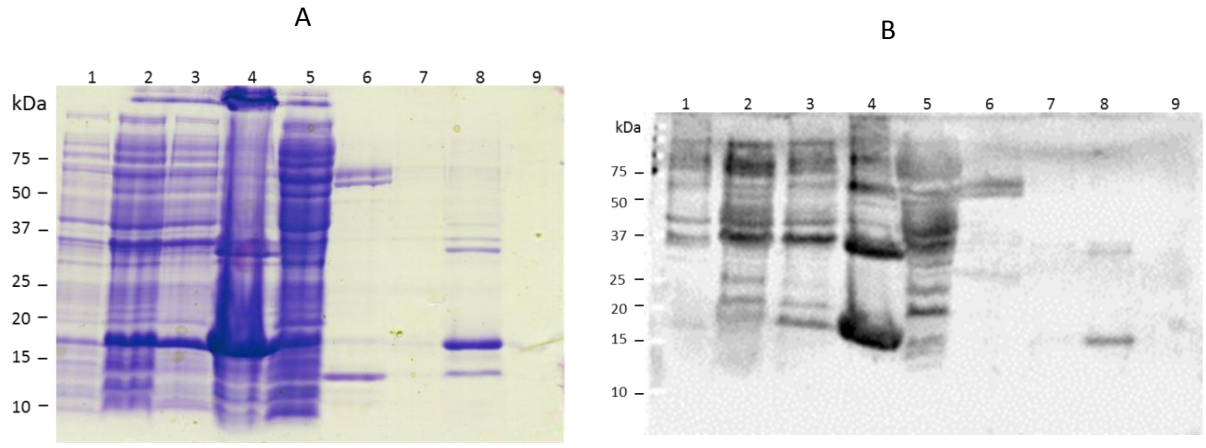

Figure 23. SDS-PAGE and Western blot analysis of recombinant CBM 120 expression and purification in HisTrap column. Overnight expression in BL21 star cells in LB medium by 0,1 mM IPTG induction, at 18 ºC. Analysis of fractions from 50, 100, 300 and 500mM imidazole elution steps. **|A.** 20µL of each sample denaturated with 2x solution buffer were applied in a 12,5% polyacrylamide gel. Gel was stained with Comassie Blue. **|B.** Western blotting. Protein´s separation by SDS-PAGE and transferred to a PVDF membrane, probed with an anti-His primary antibody (1:10 000) and detected by fluorescence using ECF as a substrate for the secondary antibody. **1**-Before IPTG induction; **2**-After 3h of IPTG induction; **3**-Insoluble fraction and **4**-Soluble fraction, after cell lysis; **5-**Flowthrough; **6**-**9** fraction of 50 mM, 100 mM, 300 mM and 500 mM imidazole steps, respectively.

From the SDS-PAGE analysis, it is observed that part of the CBM 120 protein is accumulated as inclusion bodies in accordance with the results from the small-scale expression screening. The initial fraction collected from the 300mM imidazole peak presents a major band close to 15 kDa correspondent to the CBM 120 visible with Coomassie staining. The western blot analysis confirmed the nature of the protein band at approximately 15 kDa.

A second purification step was performed for CBM 120 by analytical size exclusion chromatography in a Superdex 200 10/300 GL column. This step was used to further purify the protein and also to evaluate the protein oligomerization state.

The collected fractions from the 300 mM imidazole step were kept at 4 ºC overnight and then pooled with a final volume of 15 mL. The pooled protein sample was concentrated to a final volume of 700 µL, from where 500 µL were applied to the Superdex 200 10/300 GL column. The chromatogram (fig. 24A) shows a first peak at around the column void volume (8,16 mL) which corresponds to aggregated protein,

followed a broad peak at 10,5 mL which, according to the column calibration, should correspond to lesser aggregated forms of CBM 120. At 16 mL a lower intensity peak was detected with a elution volume compatible with dimeric form of CBM 120. The 1mL fractions collected during size-exclusion chromatography were then analyzed by SDS-PAGE followed by Coomassie staining (fig. 24B).

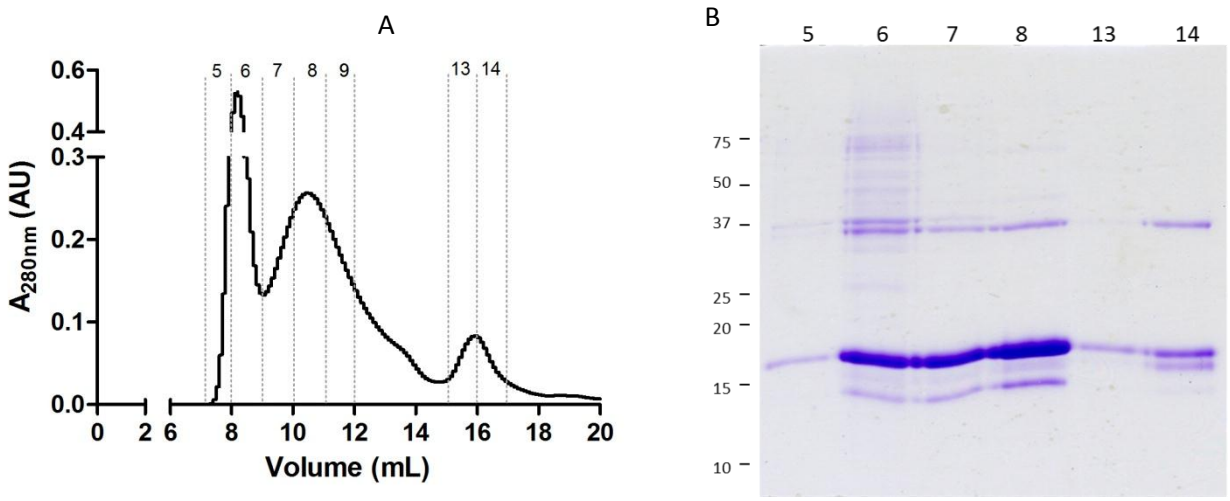

Figure 24. Analytical size-exclusion chromatography of the CBM 120 fractions eluted at the HisTrap 300mM imidazole step. |A. Chromatogram of Superdex 200 analytical purification step. 500µL the pool was applied to a Superdex 10/300 GL column equilibrated in 20mM sodium phosphate, 150mM NaCl pH 7,5. Relevant fractions are identified with the numbers. **|B**. SDS-PAGE analysis of the ASEC profile. 20µL were applied in the polyacrylamide gel, then stained with Comassie Blue.

In the SDS-PAGE it is observed in fraction 5 and 6 intense bands at the expected molecular weight for CBM 120 (15,5 kDa) however the protein presence in this fraction should be in aggregated state, as they are early eluted from the size-exclusion chromatography. Fractions 13 e 14 have shown to be mostly the CBM 120 protein, as the detection of protein dimers (band at around 37 kDa) resistant to protein denaturation have been previously described[18,54]. The protein band in sample 14 detected with a molecular weight around 15 kDa might correspond to protein truncation, protein degradation or by a non-specific removal of the His-tag by E. coli proteases.

The quantification of the results obtained along the protocol is summarized in the following purification table (Table 3).

Table 3
Purification table of recombinant CBM 120 construct expressed in soluble form.

Step	Volume (mL)	Total protein (mg)	Estimated amount of CBM 120 (mg)	Yield%	Purity%
Cell extract	22,5	429,9	147,4	100	34
Histrap	15	231,6	115,8	79	50
Superdex 200	15	2,25	1,73	1	77

Data obtained from 1L of E. coli expression culture.

The purification table shows that CBM 120 has a high tendency to aggregate, as shown by the extensive protein loss at the last chromatography step where the protein is able to be separated in to non-aggregated forms of protein and the proteins that despite in soluble form are aggregated.

To confirm if this loss of soluble CBM 120 protein occurs after the HisTrap purification or prior the purification, an analytical size exclusion chromatography was performed to evaluate the oligomerization state of the CBM 120 after the HisTrap purification and after the fraction's pool concentration. The chromatogram is shown in fig. 25.

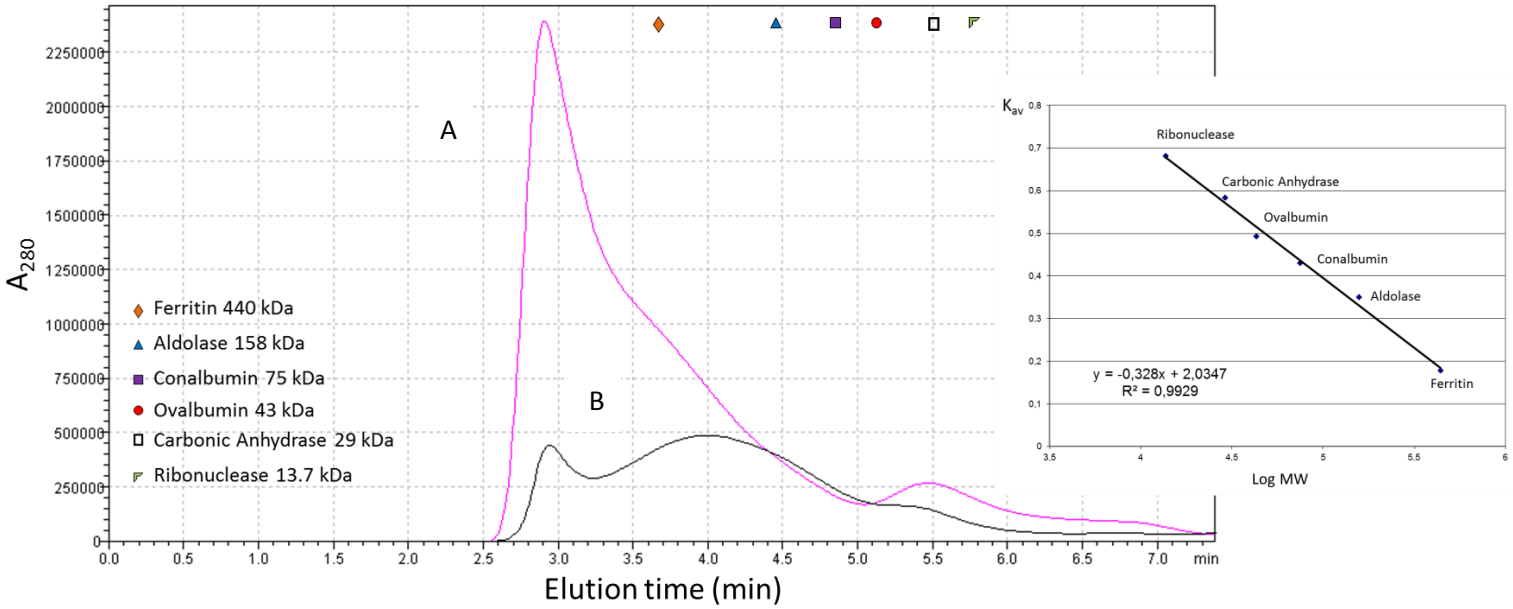

Figure 25. Analytical size-exclusion chromatography of CBM 120 first eluted fraction at 300mM imidazole step and the concentrated pool fractions of this imidazole step. 50µL of samples were applied to a Superdex 200 5/150 GL column equilibrated in 50 mM sodium phosphate pH 7.5, 150 mM NaCl and elution was carried out at 0,4 mL/min, in high pressure liquid chromatography system. Protein elution was monitored by photo diode array, and the A_{280} is displayed. Calibration proteins are outlined as symbols, positioned in the respective elution time. |A- Concentrated pool fractions of 300mM imidazole step; |B- First eluted fraction of 300 mM imidazole step. At the right, is displayed the calibration curve.

The two elution profiles, one from the sample eluted from the HisTrap with 300mM i and the same sample after being concentrated, are overlapped in fig. 25. Observing the profile A, the concentrated sample has a high content of aggregates, eluting with an elution time around 2,9 min., which corresponds to the column void volume. The less intense peak eluting at a volume of 2,2 mL (elution time 5,5 min) should correspond to the dimeric form of the CBM 120 according to the column calibration. Concerning the elution profile B that corresponds to the fraction from the Histrap directly applied to the column, it seems that some of the protein is already aggregated, right after the Histrap purification, although with a higher percentage of lower molecular weight aggregates as it is observed by the higher area of the peak right after

the void volume, indicating that protein concentration has some impact on the aggregation magnitude of the protein.

The fraction 14 eluted from the superdex 200 10/300 GL was also analyzed in analytical size exclusion chromatography to confirm its oligomerization state as a dimer using a higher resolution size exclusion chromatography column. The resulting chromatogram is displayed in fig. 26 demonstrating that the eluted protein is mostly in dimeric form as represented by the well-defined peak.

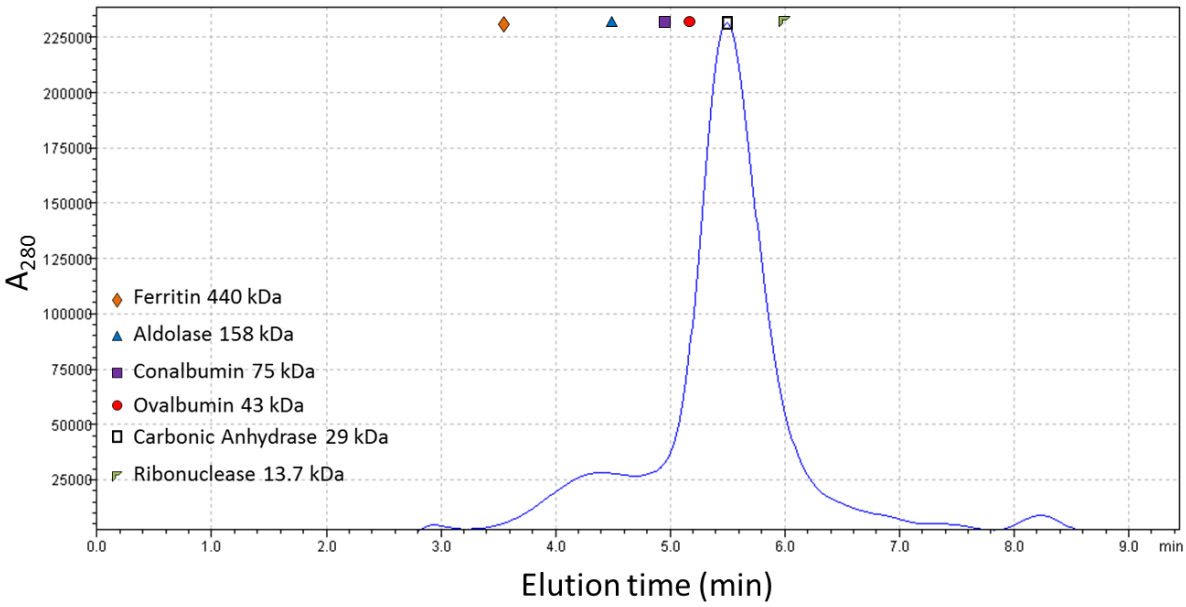

Figure 26. Analytical size-exclusion chromatography of CBM 120 fraction 14 eluted in superdex 200 10/300 GL. 50 µL of sample were injected in a Superdex 200 5 /150 GL equilibrated in 50mM sodium phosphate pH 7.5, 150mM NaCl at a flow rate of 0,4 mL/min in HPLC system. Protein elution was monitored by A_{280}. The elution time of standard proteins are denoted as symbols in the chromatogram.

CBM 120 protein is already aggregated during the HisTrap purification and the concentration of protein seems to increase its aggregation. Keeping the eluted fractions overnight with high imidazole concentration might also confer some protein instability and propensity to form aggregates. Thus, one alternative for this would be the introduction of an overnight dialysis step to remove the imidazole from the sample to prevent the long exposure with the high concentration of imidazole or the

introduction of a fast desalting step using small size-exclusion chromatography columns able to separate proteins from non-protein content. Moreover, in order to stabilize the native state of the CBM 120, additives may also be added in cell lysate such as glycerol, sorbitol, arginine, glycine or detergents. These additives might contribute to the formation of hydration shells stabilizing intramolecular protein interactions and preventing intermolecular protein interaction resulting in aggregation.

4.4. Large-scale soluble expression and purification of CBM 140 protein

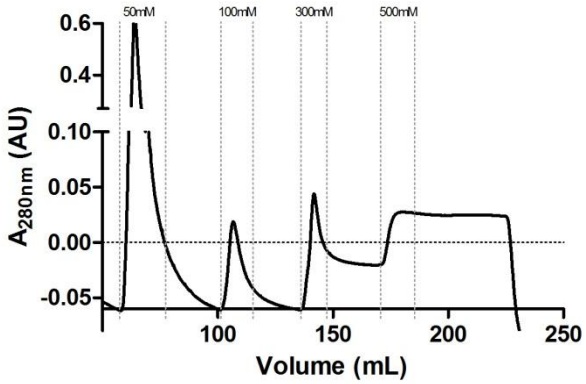

Figure 27. Chromatogram of purification of CBM 140 by immobilized metal affinity chromatography. Soluble fraction from 2L expression were loaded into a HisTrap HP 5mL column previously equilibrated in 20mM imidazole, 500mM NaCl and 20mM sodium phosphate buffer at pH 7.5. Protein elution was carried out in a four-step imidazole gradient - 50, 100, 300 and 500mM. Vertical lines outline the collected fractions in each imidazole elution step.

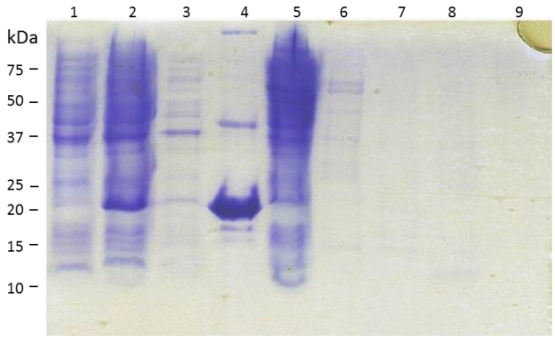

Figure 28. SDS-PAGE analysis of recombinant CBM 140 expression and purification in HisTrap column. Overnight expression in BL21 star cells in LB medium by 0,1 mM IPTG induction, at 18 ºC. Analysis of fractions from 50, 100, 300 and 500mM imidazole elution steps. 20µL of each sample were applied in a 12,5% polyacrylamide gel. **1**-Before IPTG induction; **2**-After 3h of IPTG induction; **3**-Insoluble fraction and **4**-Soluble fraction, after cell lysis; **5**-Flowthrough; **6**-**9** fraction of 50 mM, 100 mM, 300 mM and 500 mM imidazole steps, respectively.

Since some of the CBM 140 protein was detected in the soluble fraction during the small scale screening, an attempt to express in larger scale and to purify such protein in soluble was performed. From the SDS-PAGE analysis of the purification process (fig. 28) almost all the protein was accumulated as inclusion bodies, which resulted in no protein being purified by IMAC, at least within the levels detected by the coomassie stained SDS-PAGE analysis.

Taking this result into account, we have attempted the expression of CBM 140 and CBM 159 in the form of inclusion bodies, followed by a rapid dilution refolding step, following the protocol described by Castanheira et al. The oligomerization/aggregation state of refolded protein was evaluated by analytical size exclusion chromatography (Superdex 200 5/150 GL) with results revealing that both proteins were completely aggregated (data not shown).

4.5. Large-scale overexpression and purification of Malin

Regarding malin, two constructs were available in the laboratory as mentioned in the section of materials and methods (table 4). The malin_pET-GST, kindly provided by Dr. Matthew Gentry (University of Kentucky, USA), coding for GST fusing tag at N-terminal of malin and a C-terminal 6×His tag and malin_pSKB3, coding a 6×His tag at malin N-terminal. We have started by expressing malin_pSKB3 construct in 4L of LB medium. The expression was carried out overnight at 18 ºC after induction with 0,1mM of IPTG. The cells were harvested and ressuspended in binding buffer, and purification was performed by a HisTrap column, being the respective chromatogram displayed in figure 29 A. It is observed a major peak at 300 mM of imidazole step. Fractions were collected during expression and also along the protein elution then analyzed by SDS-PAGE and western blot as demonstrated in fig. 29 B and C.

From the SDS-PAGE analysis it is observed that low levels of protein were purified under the IMAC purification (lanes 6-9), being malin only detected by western blot (band at around 42 kDa in lanes 6-8) without correspondence to any band detected by coomassie stained SDS-PAGE. It is visible that, in resemblance with the CBM results, almost all the protein was retained in the insoluble fraction (lane 4). The detection of ≅ 25 and 15 kDa bands in western blot may represent products from degradation of malin by E. coli proteases which are already visible during the expression step. Despite E. coli BL21 star strain is deficient of *lon* and *Omp*T proteases, other proteases are present and may degrade the target protein.

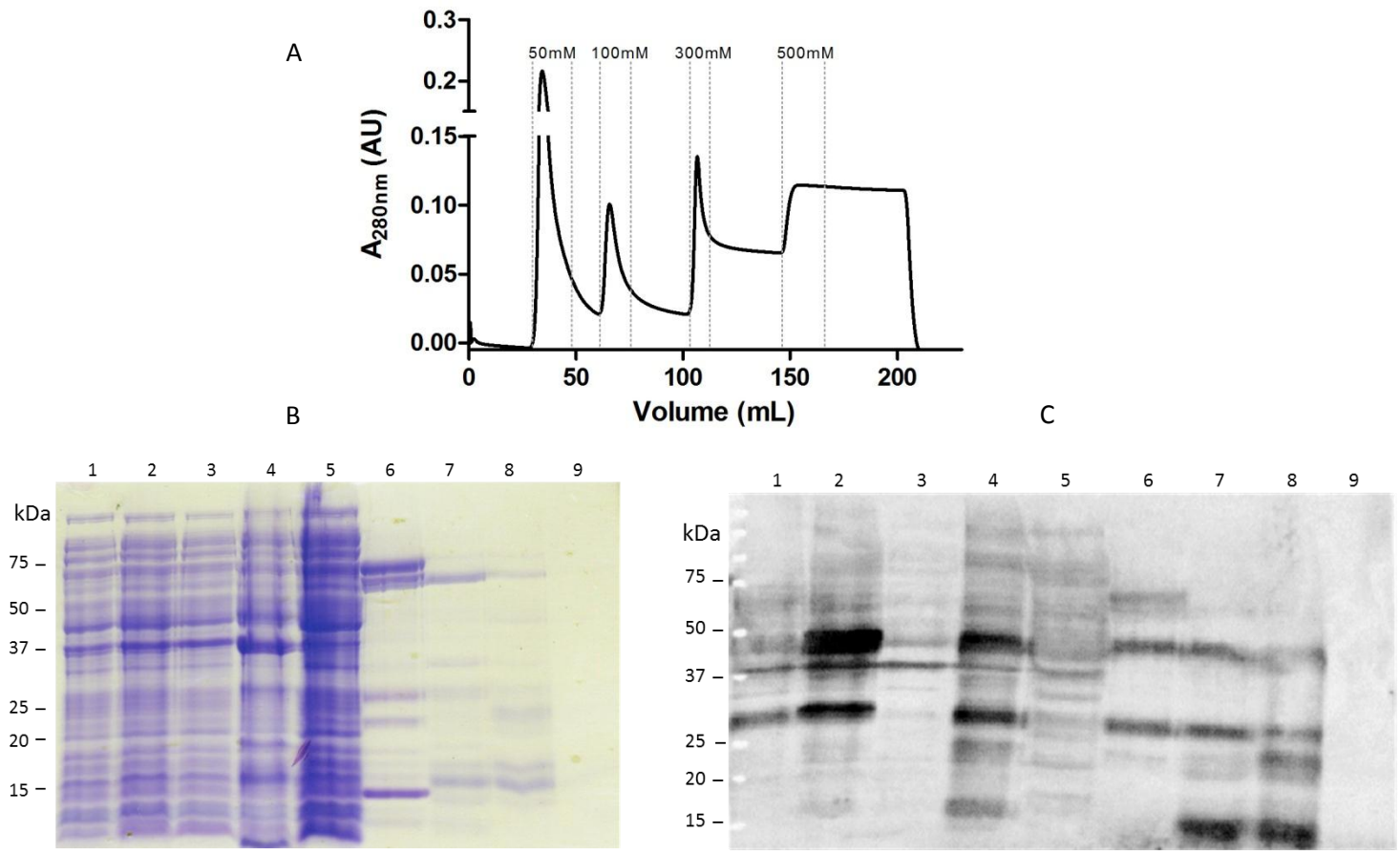

Figure 29. Purification of His_6-Malin expressed in E. Coli BL21 Star by immobilized metal ion affinity chromatography (expression volume-4L). |A. Chromatogram obtained in HisTrap HP 5mL purification step. Protein elution was carried out in a four-step imidazole gradient-50, 100, 300 and 500mM-in the same buffer, Vertical lines outline the collected fractions in each imidazole elution step. **|B**. SDS-PAGE analysis of expression and purification profile20µL of sample was loaded in 12.5% polyacrylamide gels **|C.** Western blotting. Protein´s separation by SDS-PAGE and transferred to a PVDF membrane, probed with an anti-His primary antibody (1:10 000) and detected by fluorescence using ECF as a substrate for the secondary antibody. **1**-Before IPTG induction; **2**-After 3h of IPTG induction; **3**-Soluble fraction and **4**-Insoluble fraction, after cell lysis; **5-**Flowthrough; **6-9** fraction of 50 mM, 100 mM, 300 mM and 500 mM imidazole steps, respectively.

The other construct, malin_pET-GST was also expressed in E. coli and purified by the GST-tag, however we have observed the same tendency of the protein to accumulate as inclusion bodies and low levels of protein expression, without a clear accumulation of protein visible by coomassie stained SDS-PAGE, were obtained.

4.6. Small-scale overexpression screening of recombinant Malin

The incorrect or absence of disulfide bonds formation can become a drawback in obtaining soluble and active recombinant protein. In the case of malin it was observed inclusion bodies formation in the previous results. Malin contains 17 cysteine residues along the protein sequence and the E coli cell cytoplasm is a reducing environment difficult for disulfide bond formation. This process only takes place in the periplasm through the Dsb system[56]. This limitation may contribute to the low levels of soluble purified malin. Based on this, screenings were performed for malin_pSKB3 and malin_ pET-GST with E coli BL21 star expressing Erv1p sulfhydryl oxidase and DsbC disulfide isomerase or PDI protein disulfide isomerase[57]. The bacteria is previously transformed with pBAD promotor vectors containing the coding genes for Erv1p and DsbC or Erv1p and PDI. The Erv1p is a sulfhydryl oxidase from the inter-membrane mitochondrial space of S. cerevisiae. This enzyme is a natural catalyst of *de novo* disulfide bond formation. The DsbC is an E Coli. periplasmic isomerase that catalysis disulfide bond isomerization and PDI is a protein disulfide isomerase present in the endoplasmic reticulum of eukaryotes and also catalyzes the rearrangement of disulfide bonds within proteins as they fold. The use of these enzymes showed efficient production of eukaryotic proteins with multiple disulfide bonds in the cytoplasm of E. coli[57]. Taking this into account, double screening was performed using E coli BL21 star, E coli BL21 star pre-expressing Erv1p and DsbC, and E coli BL21 star pre-expressing Erv1p and PDI. The screenings were performed in 50 mL of cell culture, with 0,5% arabinose being added to induce the pre-expression of the folding inducing enzymes. Then malin expression was induced by 0.1 mM of IPTG and expression allowed to occur overnight at 18 ºC. Expression was analyzed by SDS-PAGE gel and Anti-His Western blotting for malin_pSKB3 (fig. 30) and malin_pET-GST (fig. 31).

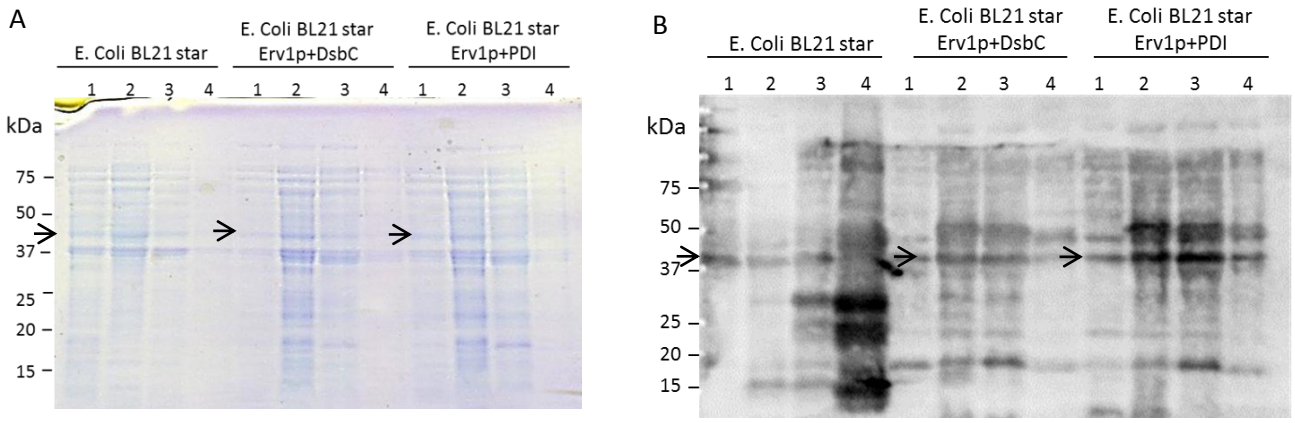

Figure 30. Small-scale overexpression screening of recombinant His_6-Malin in E. Coli BL21 star and E. Coli BL21 star pre-expressing Erv1p oxidase and isomerases DsbC or PDI. Expression was carried out using LB medium at 18ºC, for overnight growth. **|A**. SDS-PAGE. **|B**. Western blotting. All samples were denaturated with 2x loading solution and 20µL were applied in 12.5% polyacrylamide gels. For WB, proteins were transferred to a PVDF membrane, probed with an anti-His primary antibody (1:10 000) and detected by chemiluminescense using ECF as a substrate of the secondary antibody. **1**- Before IPTG induction, **2**- After IPTG induction, **3**- Soluble fraction and **4**- Insoluble fraction after cell lysis.

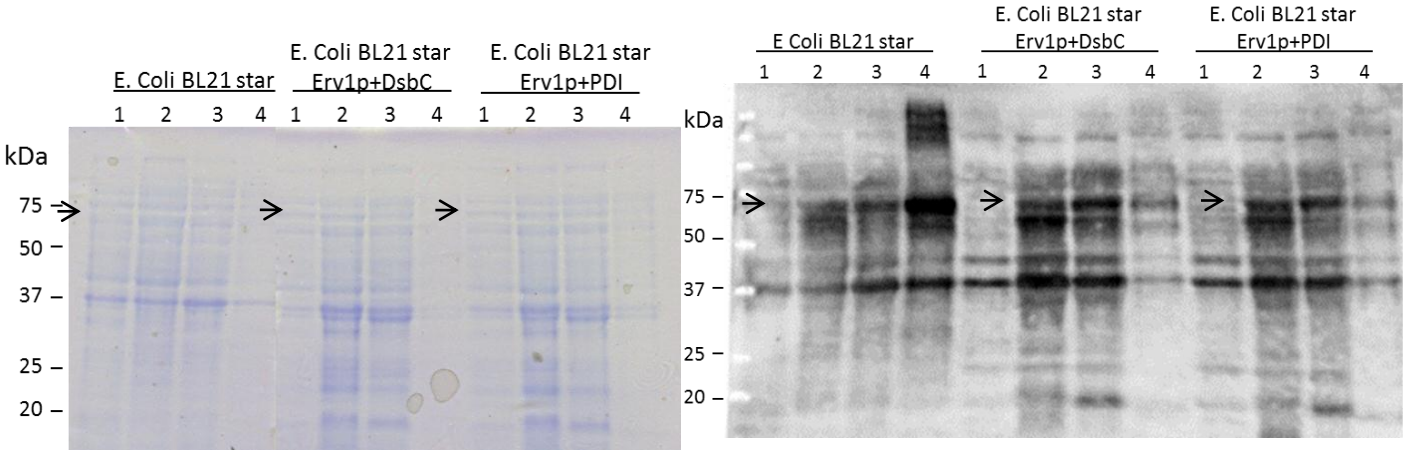

Figure 31. Small-scale overexpression screening of recombinant T7-GST-Malin-His_6 in E. Coli BL21 star and E. Coli BL21 star plus Erv1p oxidase and isomerases DsbC or PDI. Expression was carried out using LB medium at 18ºC, for overnight growth. **|A**. SDS-PAGE. **|B**. Western blotting. 20µL of sample were applied in 12.5% polyacrylamide gels. For WB, proteins were transferred to a PVDF membrane, probed with an anti-His primary antibody (1:10 000) and detected by fluorescence using ECF as a substrate of the secondary antibody. **1**- Before IPTG induction, **2**- After IPTG induction, **3**- Soluble fraction and **4**- Insoluble fraction after cell lysis.

Concerning the SDS-PAGE and western blot analysis of His_6-Malin, more protein seems to be present in the soluble form, when co-expressed with Erv1p + PDI (fig. 30 lane 3 Erv1p + PDI condition) in comparison with the expression conditions or no pre-expression of "foldases" was performed (fig. 30 BL21 star condition – lane 3), nevertheless the levels of protein expression remained low. Relatively to the GST-Malin-His_6, it is observed mainly in the WB the presence of malin near 75kDa. Expression in E Coli BL21 star, the low amount of the protein expressed was aggregated as observed in the insoluble fraction (lane 4). In the E Coli (Erv1p+ DsbC) a band detected by the WB at 75 kDa in the soluble fraction demonstrates the increase of protein in the soluble form and in E Coli (Erv1p+ PDI) the same is observed however the band is not so intense. A band pattern is also present at 37 kDa in the WB. Since the antibody used is against His and malin's molecular weight is about 42 kDa, these bands should correspond to His-malin without the T7-GST tag. This was already observed in the previous expressions and purifications. Here, is shown an increase in soluble protein for the E coli strains with the oxidase and isomerases being more evident in the presence of the Erv1p with DsbC. Taking this into account, the E. coli strain selected for the scale-up expression of malin_PET-GST was Coli (Erv1p+ DsbC).

4.7 Large-scale overexpression and GST-Malin-His6 constructs in the presence of disulfide bond catalysts.

The next attempt was to express 2L of GST-Malin-His_6 in the presence of the plasmid coding for Erv1p and DsbC. The pre-expression of Erv1p and PDI was induced by 0,5% arabinose and after that, 0,1 mM of IPTG was added to induce malin expression that was allowed to occur overnight at 18ºC. The purification was carried out by IMAC. Fractions from expression and from purification were analyzed in SDS-PAGE and WB (fig. 32).

This results demonstrate that at the presence of the oxidase and foldase, a considerable increase in soluble form of malin is obtained at soluble fraction (fig. 32 B, lane 3), however the increase in the soluble protein being purified could again only observed by the western blot analysis (fig. 32 C), with no corresponding visible band being detected by the SDS-PAGE analysis (fig. 32 B).

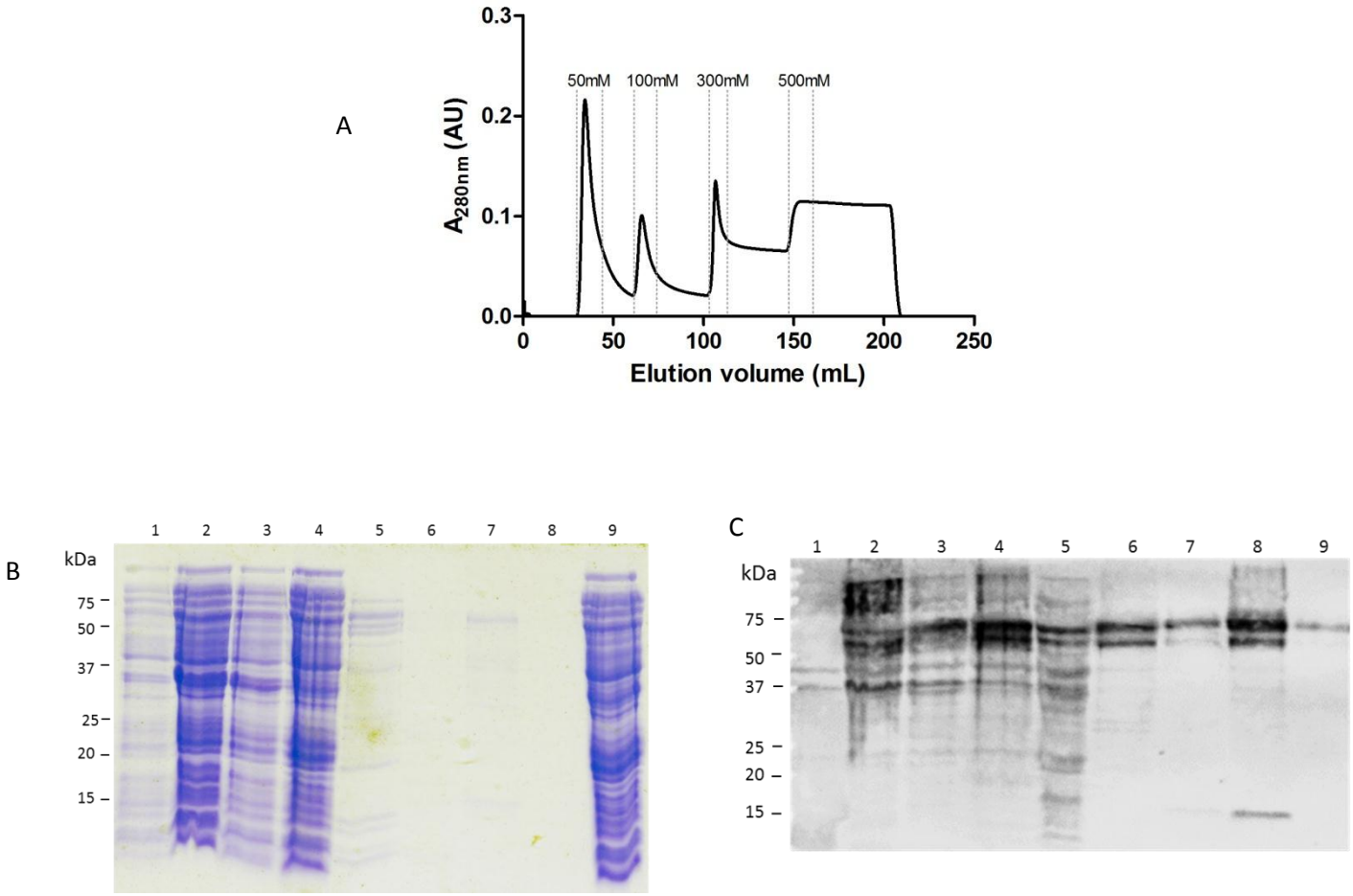

Figure 32. Purification of GST-Malin-His_6 expressed in E. Coli BL21 Star (Erv1p + DsbC) by immobilized metal ion affinity chromatography (expression volume-2L). |A. Chromatogram obtained in HisTrap HP 5mL purification step. Vertical lines outline the collected fractions in each imidazole elution step. **|B**. SDS-PAGE analysis of expression and purification profile. 20µL of sample was loaded in 12.5% polyacrylamide gels **1**- Before IPTG induction, **2**- After IPTG induction (after 3h), **3**- Soluble fraction and **4**- Insoluble fraction after cell lysis, **5**- fraction of 50mM imidazole step, **6-** fraction of 100mM imidazole step, **7**- fraction of 300mM imidazole step, **8**- fraction of 500mM imidazole step, **9**- Flow through. **|C.** Western blotting. Protein´s separation by SDS-PAGE and transferred to a PVDF membrane, probed with an anti-His primary antibody (1:10 000) and detected by chemiluminescense using ECF as a substrate for the secondary antibody. **1**-Before IPTG induction; **2**-After 3h of IPTG induction; **3**-Soluble fraction and **4**-Insoluble fraction, after cell lysis; **5-**Flowthrough; **6-9** fraction of 50 mM, 100 mM, 300 mM and 500 mM imidazole steps, respectively.

The fractions collected in the 300mM imidazole step were pooled, dialyzed overnight and concentrated to a final volume of 700µL. The protein was monitored relatively to its oligomerization state by analytical size exclusion chromatography, confirming very low yields of expressed and purified malin and a major presence of the GST tag (data not shown).

A possible reason for these results is the divergent codon usage between humans and E. coli, since recombinant protein expression in heterologous systems can be impaired by different codon usage because the concentrations of host tRNAs, in this case E coli, are insufficient for the less-used codons to optimally translate the mRNA. Indeed, it is visible the difference of the codon quality for E. coli from the original to the codon optimized malin coding sequence (fig. 33).

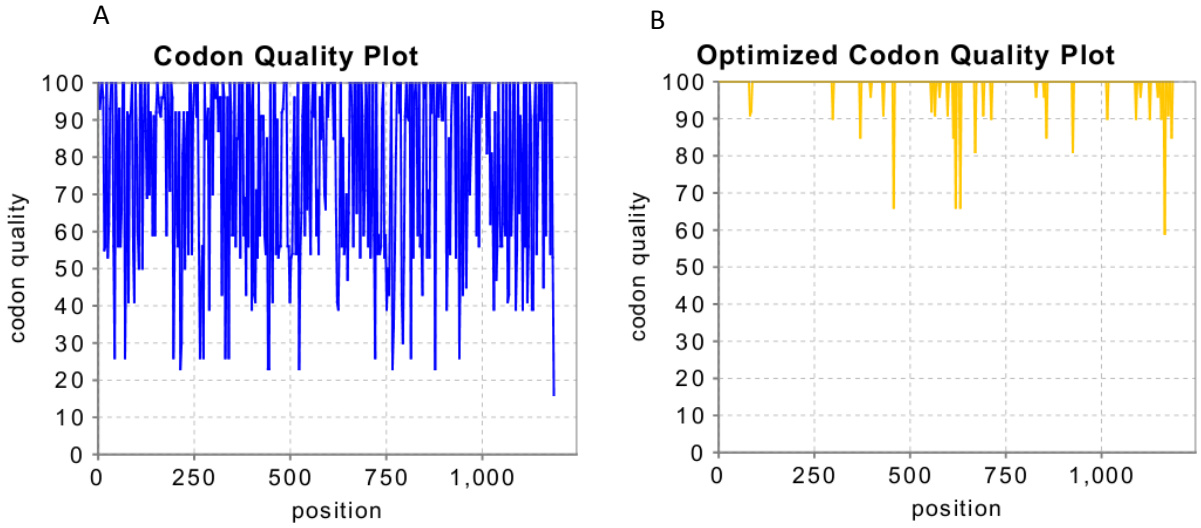

Figure 33. Plots showing the quality of used codon of the original malin sequence (A) and the optimized malin sequence (B). The X axis represents the nucleotide position.

5. Conclusions and future perspectives

The major goal of the developed work was to optimize the expression and purification of the laforin derived constructs and malin not only in sufficient amounts but also in a soluble and stable form suitable for later interaction characterization.

The generation of the laforin derived constructs using the original laforin coding sequence as the template gave amplification problems for almost all constructs and unreadable sequencing results of the cloned DNA sequences due to the high GC content present at the first half of laforin sequence. This drawback was successfully circumvented by using an optimized GC content and codon optimized laforin coding sequence with clearly improvement in codon quality and decreased GC levels. This sequence optimization besides promoting efficiency on the amplification and cloning process it also contributes to optimize the expression levels.

From the expression screening of the CBM constructs, CBM 120, 140 and 159 along with DSP 111 and 130, it's concluded that an increase of the sequence from the inter-domain region contributes to increasing in protein accumulation in the form of inclusion bodies. Published data supports this evidence that in the presence of soluble domains, the addition of extra amino acid sequence tends to affect negatively the solubility and expression yields of the domains.

Expressing CBM 120 construct in large scale in E. coli Bl21 star it was visible the presence of soluble protein however, a high amount was accumulated in inclusion bodies as expected from the screening analysis. CBM 120 was further purified and concentrated with a good final yield of 1,73 mg per liter of expression however, the protein concentration promoted more aggregation of CBM 120 . Attempts in refolding of the CBM 120 and CBM 140 proteins from inclusion bodies were unsuccessful, since protein was aggregated.

Malin expression was followed, using a previously described expression protocol, however protein was poorly expressed in E. coli, being only detected by western-blot, being the small amounts expressed, accumulated as inclusion bodies. Attempts of expressing malin in the presence of sulfhydryl oxidase and isomerases as way of prevent insoluble aggregation and proper folding of malin were marginally successful

since soluble protein levels were higher, but were still insufficient for protein to be purified. The possible reason is the expression of a non-codon optimized malin sequence for E. coli. Recombinant protein expression in heterologous systems can be impaired by different codon usage because the concentrations of host tRNAs, in this case E coli, are insufficient for the less-used codons to optimally translate the mRNA[58]. The final result is no protein expressed or partial expression of the first amino acids, including tags if present on the N-terminal giving final products of incomplete protein synthesis.

Considering this, further work including obtaining an optimized malin coding sequence for expression in E. coli; testing fusion tags in laforin derived constructs which increase solubility of target proteins, such as MBP, Thioredoxin, Nus A or SUMO; the co-expression of chaperones to improve the correct folding and thus solubility of the proteins; the use of pET Duet vector to co-express each laforin derived construct and malin allowing in vivo interaction and final purification of the complex, is proposed.

References

1. Roach, P. J. (2002). "Glycogen and its metabolism." Current molecular medicine **2**(2): 101-120.

2. Tagliabracci, V. S., C. Heiss, C. Karthik, C. J. Contreras, J. Glushka, M. Ishihara, P. Azadi, T. D. Hurley, A. A. DePaoli-Roach and P. J. Roach (2011). "Phosphate Incorporation during Glycogen Synthesis and Lafora Disease." Cell Metabolism **13**(3): 274-282.

3. Vernia, S., M. Carmen Solaz-Fuster, J. Vicente Gimeno-Alcaniz, T. Rubio, L. Garcia-Haro, M. Foretz, S. Rodriguez de Cordoba and P. Sanz (2009). "AMP-activated Protein Kinase Phosphorylates R5/PTG, the Glycogen Targeting Subunit of the R5/PTG-Protein Phosphatase 1 Holoenzyme, and Accelerates Its Down-regulation by the Laforin-Malin Complex." Journal of Biological Chemistry **284**(13): 8247-8255.

4. Fong, N. M., T. C. Jensen, A. S. Shah, N. N. Parekh, A. R. Saltiel and M. J. Brady (2000). "Identification of binding sites on protein targeting to glycogen for enzymes of glycogen metabolism." Journal of Biological Chemistry **275**(45): 35034-35039.

5. Minassian, B. A., J. R. Lee, J. A. Herbrick, J. Huizenga, S. Soder, A. J. Mungall, I. Dunham, R. Gardner, C. G. Fong, S. Carpenter, L. Jardim, P. Satishchandra, E. Andermann, O. C. Snead, I. Lopes-Cendes, L. C. Tsui, A. V. Delgado-Escueta, G. A. Rouleau and S. W. Scherer (1998). "Mutations in a gene encoding a novel protein tyrosine phosphatase cause progressive myoclonus epilepsy." Nature Genetics **20**(2): 171-174.

6. Serratosa, J. M., P. Gomez-Garre, M. E. Gallardo, B. Anta, D. B. V. de Bernabe, D. Lindhout, P. B. Augustijn, C. A. Tassinari, R. Michelucci, A. Malafosse, M. Topcu, D. Grid, C. Dravet, S. F. Berkovic and S. R. de Cordoba (1999). "A novel protein tyrosine phosphatase gene is mutated in progressive myoclonus epilepsy of the Lafora type (EPM2)." Human Molecular Genetics **8**(2): 345-352.

7. Dukhande, V. V., D. M. Rogers, C. Roma-Mateo, J. Donderis, A. Marina, A. O. Taylor, P. Sanz and M. S. Gentry (2011). "Laforin, a dual specificity phosphatase involved in Lafora disease, is present mainly as monomeric form with full phosphatase activity." PloS one **6**(8): e24040.

8. Ganesh, S., N. Tsurutani, T. Suzuki, Y. Hoshii, T. Ishihara, A. V. Delgado-Escueta and K. Yamakawa (2004). "The carbohydrate-binding domain of Lafora disease protein targets Lafora polyglucosan bodies." Biochemical and Biophysical Research Communications **313**(4): 1101-1109.

9. Dubey, D. and S. Ganesh (2008). "Modulation of functional properties of laforin phosphatase by alternative splicing reveals a novel mechanism for the EPM2A gene in Lafora progressive myoclonus epilepsy." Human Molecular Genetics **17**(19): 3010-3020.

10. Worby, C. A., M. S. Gentry and J. E. Dixon (2006). "Laforin, a dual specificity phosphatase that dephosphorylates complex carbohydrates." Journal of Biological Chemistry **281**(41): 30412-30418.

11. Roma-Mateo, C., M. del Carmen Solaz-Fuster, J. Vicente Gimeno-Alcaniz, V. V. Dukhande, J. Donderis, C. A. Worby, A. Marina, O. Criado, A. Koller, S. Rodriguez De Cordoba, M. S. Gentry and P. Sanz (2011). "Laforin, a dual-specificity phosphatase involved in Lafora disease, is phosphorylated at Ser(25) by AMP-activated protein kinase." Biochemical Journal **439**: 265-275.

12. Puri, R., T. Suzuki, K. Yamakawa and S. Ganesh (2009). "Hyperphosphorylation and Aggregation of Tau in Laforin-deficient Mice, an Animal Model for Lafora Disease." Journal of Biological Chemistry **284**(34): 22657-22663.

13. Ganesh, S., R. Puri, S. Singh, S. Mittal and D. Dubey (2006). "Recent advances in the molecular basis of Lafora's progressive myoclonus epilepsy." Journal of Human Genetics **51**(1): 1-8.

14. Wang, J. Y., J. A. Stuckey, M. J. Wishart and J. E. Dixon (2002). "A unique carbohydrate binding domain targets the Lafora disease phosphatase to glycogen." Journal of Biological Chemistry **277**(4): 2377-2380.

15. Tonks, N. K. (2006). "Protein tyrosine phosphatases: from genes, to function, to disease." Nature Reviews Molecular Cell Biology **7**(11): 833-846.

16. Liu, Y., Y. Wang, C. Wu, Y. Liu and P. Zheng (2006). "Dimerization of Laforin is required for its optimal phosphatase activity, regulation of GSK3 beta phosphorylation, and Wnt signaling." Journal of Biological Chemistry **281**(46): 34768-34774.

17. Fernandez-Sanchez, M. E., O. Criado-Garcia, K. E. Heath, B. Garcia-Fojeda, I. Medrano-Fernandez, P. Gomez-Garre, P. Sanz, J. M. Serratosa and S. R. de Cordoba (2003). "Laforin, the dual-phosphatase responsible for Lafora disease, interacts with R5 (PTG), a regulatory subunit of protein phosphatase-1 that enhances glycogen accumulation." Human Molecular Genetics **12**(23): 3161-3171.

18. Moreira, S., P. Castanheira, M. Casal, C. Faro and M. Gama (2010). "Expression of the functional carbohydrate-binding module (CBM) of human laforin." Protein Expression and Purification **74**(2): 169-174.

19. Chan, E. M., E. J. Young, L. Ianzano, I. Munteanu, X. C. Zhao, C. C. Christopoulos, G. Avanzini, M. Elia, C. A. Ackerley, N. J. Jovic, S. Bohlega, E. Andermann, G. A. Rouleau, A. V. Delgado-Escueta, B. A. Minassian and S. W. Scherer (2003). "Mutations in NHLRC1 cause progressive myoclonus epilepsy." Nature Genetics **35**(2): 125-127.

20. Deshaies, R. J. and C. A. P. Joazeiro (2009). "RING Domain E3 Ubiquitin Ligases." Annual Review of Biochemistry **78**: 399-434.

21. Freemont, P. S. (2000). "Ubiquitination: RING for destruction?" Current Biology **10**(2): R84-R87.

22. Edwards, T. A., B. D. Wilkinson, R. P. Wharton and A. K. Aggarwal (2003). "Model of the Brain Tumor-Pumilio translation repressor complex." Genes & Development **17**(20): 2508-2513.

23. Slack, F. J. and G. Ruvkun (1998). "A novel repeat domain that is often associated with RING finger and B-box motifs." Trends in Biochemical Sciences **23**(12): 474-475.

24. Gentry, M. S., C. A. Worby and J. E. Dixon (2005). "Insights into Lafora disease: Malin is an E3 ubiquitin ligase that ubiquitinates and promotes the degradation of laforin." Proceedings of the National Academy of Sciences of the United States of America **102**(24): 8501-8506.

25. Worby, C. A., M. S. Gentry and J. E. Dixon (2008). "Malin decreases glycogen accumulation by promoting the degradation of protein targeting to glycogen (PTG)." Journal of Biological Chemistry **283**(7): 4069-4076.

26. Tagliabracci, V. S., J. M. Girard, D. Segvich, C. Meyer, J. Turnbull, X. Zhao, B. A. Minassian, A. A. DePaoli-Roach and P. J. Roach (2008). "Abnormal Metabolism of Glycogen Phosphate as a Cause for Lafora Disease." Journal of Biological Chemistry **283**(49): 33816-33825.

27. Lohi, H., L. Ianzano, X. C. Zhao, E. M. Chan, J. Turnbull, S. W. Scherer, C. A. Ackerley and B. A. Minassian (2005). "Novel glycogen synthase kinase 3 and ubiquitination pathways in progressive myoclonus epilepsy." Human Molecular Genetics **14**(18): 2727-2736.

28. Vilchez, D., S. Ros, D. Cifuentes, L. Pujadas, J. Valles, B. Garcia-Fojeda, O. Criado-Garcia, E. Fernandez-Sanchez, I. Medrano-Fernandez, J. Dominguez, M. Garcia-Rocha, E. Soriano, S. R. De Cordoba and J. J. Guinovart (2007). "Mechanism suppressing glycogen synthesis in neurons and its demise in progressive myoclonus epilepsy." Nature Neuroscience **10**(11): 1407-1413.

29. Carmen Solaz-Fuster, M., J. Vicente Gimeno-Alcaniz, S. Ros, M. Elena Fernandez-Sanchez, B. Garcia-Fojeda, O. Criado Garcia, D. Vilchez, J. Dominguez, M. Garcia-Rocha, M. Sanchez-Piris, C. Aguado, E. Knecht, J. Serratosa, J. Josep Guinovart, P. Sanz and S. Rodriguez de Cordoba (2008). "Regulation of glycogen synthesis by the laforin-malin complex is modulated by the AMP-activated protein kinase pathway." Human Molecular Genetics **17**(5): 667-678.

30. Cheng, A., M. Zhang, M. S. Gentry, C. A. Worby, J. E. Dixon and A. R. Saltiel (2007). "A role for AGL ubiquitination in the glycogen storage disorders of Lafora and Cori's disease." Genes & Development **21**(19): 2399-2409.

31. Sharma, J., S. N. R. Rao, S. K. Shankar, P. Satishchandra and N. R. Jana (2011). "Lafora disease ubiquitin ligase malin promotes proteasomal degradation of neuronatin and regulates glycogen synthesis." Neurobiology of Disease **44**(1): 133-141.

32. Delgado-Escueta, A. V. (2007). "Advances in Lafora progressive myoclonus epilepsy." Current Neurology and Neuroscience Reports **7**(5): 428-433.

33. Singh, S. and S. Ganesh (2009). "Lafora Progressive Myoclonus Epilepsy: A Meta-analysis of Reported Mutations in the First Decade following the Discovery of the EPM2A and NHLRC1 Genes." Human Mutation **30**(5): 715-723.

34. Minassian, B. A. (2001). "Lafora's disease: Towards a clinical, pathologic, and molecular synthesis." Pediatric Neurology **25**(1): 21-29.

35. Striano, P., F. Zara, J. Turnbull, J.-M. Girard, C. A. Ackerley, M. Cervasio, G. De Rosa, M. L. Del Basso-De Caro, S. Striano and B. A. Minassian (2008). "Typical progression of myoclonic epilepsy of the Lafora type: a case report." Nature Clinical Practice Neurology **4**(2): 106-111.

36. Andrade, D. M., J. Turnbull and B. A. Minassian (2007). "Lafora disease, seizures and sugars." Acta myologica : myopathies and cardiomyopathies : official journal of the Mediterranean Society of Myology / edited by the Gaetano Conte Academy for the study of striated muscle diseases **26**(1): 83-86.

37. Ganesh, S., A. V. Delgado-Escueta, T. Sakamoto, M. R. Avila, J. Machado-Salas, Y. Hoshii, T. Akagi, H. Gomi, T. Suzuki, K. Amano, K. L. Agarwala, Y. Hasegawa, D. S. Bai, T. Ishihara, T. Hashikawa, S. Itohara, E. M. Cornford, H. Niki and K. Yamakawa (2002). "Targeted disruption of the Epm2a gene causes formation of Lafora inclusion bodies, neurodegeneration, ataxia, myoclonus epilepsy and impaired behavioral response in mice." Human Molecular Genetics **11**(11): 1251-1262.

38. Worby, C. A., M. S. Gentry and J. E. Dixon (2006). "Laforin, a dual specificity phosphatase that dephosphorylates complex carbohydrates." Journal of Biological Chemistry **281**(41): 30412-30418.

39. Girard, J.-M., K. H. D. Le and F. Lederer (2006). "Molecular characterization of laforin, a dual-specificity protein phosphatase implicated in Lafora disease." Biochimie **88**(12): 1961-1971

40. Chan, E. M., E. J. Young, L. Ianzano, I. Munteanu, X. C. Zhao, C. C. Christopoulos, G. Avanzini, M. Elia, C. A. Ackerley, N. J. Jovic, S. Bohlega, E. Andermann, G. A. Rouleau, A. V. Delgado-Escueta, B. A. Minassian and S. W. Scherer (2003). "Mutations in NHLRC1 cause progressive myoclonus epilepsy." Nature Genetics **35**(2): 125-127.d

41. Wang, W., G. E. Parker, A. V. Skurat, N. Raben, A. A. DePaoli-Roach and P. J. Roach (2006). "Relationship between glycogen accumulation and the laforin dual specificity phosphatase." Biochemical and Biophysical Research Communications **350**(3): 588-592.

42. Tagliabracci, V. S., J. Turnbull, W. Wang, J.-M. Girard, X. Zhao, A. V. Skurat, A. V. Delgado-Escueta, B. A. Minassian, A. A. DePaoli-Roach and P. J. Roach (2007). "Laforin is a glycogen phosphatase, deficiency of which leads to elevated phosphorylation of glycogen in vivo." Proceedings of the National Academy of Sciences of the United States of America **104**(49): 19262-19266.

43. Valles-Ortega, J., J. Duran, M. Garcia-Rocha, C. Bosch, I. Saez, L. Pujadas, A. Serafin, X. Canas, E. Soriano, J. M. Delgado-Garcia, A. Gruart and J. J. Guinovart (2011). "Neurodegeneration and functional impairments associated with glycogen synthase accumulation in a mouse model of Lafora disease." EMBO molecular medicine **3**(11): 667-681.

44. Garyali, P., P. Siwach, P. K. Singh, R. Puri, S. Mittal, S. Sengupta, R. Parihar and S. Ganesh (2009). "The malin-laforin complex suppresses the cellular toxicity of misfolded proteins by promoting their degradation through the ubiquitin-proteasome system." Human Molecular Genetics **18**(4): 688-700.

45. Mittal, S., D. Dubey, K. Yamakawa and S. Ganesh (2007). "Lafora disease proteins malin and laforin are recruited to aggresomes in response to proteasomal impairment." Human Molecular Genetics **16**(7): 753-762.

46. Aguado, C., S. Sarkar, V. I. Korolchuk, O. Criado, S. Vernia, P. Boya, P. Sanz, S. Rodriguez de Cordoba, E. Knecht and D. C. Rubinsztein (2010). "Laforin, the most common protein mutated in Lafora disease, regulates autophagy." Human Molecular Genetics **19**(14): 2867-2876.

47. Criado, O., C. Aguado, J. Gayarre, L. Duran-Trio, A. M. Garcia-Cabrero, S. Vernia, B. San Millan, M. Heredia, C. Roma-Mateo, S. Mouron, L. Juana-Lopez, M. Dominguez, C. Navarro, J. M. Serratosa, M. Sanchez, P. Sanz, P. Bovolenta, E. Knecht and S. Rodriguez de Cordoba (2012). "Lafora bodies and neurological defects in malin-deficient mice correlate with impaired autophagy." Human Molecular Genetics **21**(7): 1521-1533.

48. Knecht, E., C. Aguado, S. Sarkar, V. I. Korolchuk, O. Criado-Garcia, S. Vernia, P. Boya, P. Sanz, S. Rodriguez de Cordoba and D. C. Rubinsztein (2010). "Impaired autophagy in Lafora disease." Autophagy **6**(7): 991-993.

49. Puri, R. and S. Ganesh (2010). "Laforin in autophagy A possible link between carbohydrate and protein in Lafora disease?" Autophagy **6**(8): 1229-1231.

50. Chakrabarti, R. and C. E. Schutt (2001). "The enhancement of PCR amplification by low molecular-weight sulfones." Gene **274**(1-2): 293-298.

51. Frackman, S., G. Kobs, D. Simpson and D. Storts (1998). Betaine and DMSO: Enhancing Agents for PCR Promega Notes**:** 27 - 30.

52. Roux, K. H. (1995). "OPTIMIZATION AND TROUBLESHOOTING IN PCR." Pcr-Methods and Applications **4**(5): S185-S194.

53. An, Y., H. Yumerefendi, P. J. Mas, A. Chesneau and D. J. Hart (2011). "ORF-selector ESPRIT: A second generation library screen for soluble protein expression employing precise open reading frame selection." Journal of Structural Biology **175**(2): 189-197.

54. Liu, Y., Y. Wang, C. Wu, Y. Liu and P. Zheng (2006). "Dimerization of Laforin is required for its optimal phosphatase activity, regulation of GSK3 beta phosphorylation, and Wnt signaling." Journal of Biological Chemistry **281**(46): 34768-34774.

55. Gentry, M. S., C. A. Worby and J. E. Dixon (2005). "Insights into Lafora disease: Malin is an E3 ubiquitin ligase that ubiquitinates and promotes the degradation of laforin." Proceedings of the National Academy of Sciences of the United States of America **102**(24): 8501-8506.

56. Prinz, W. A., F. Aslund, A. Holmgren and J. Beckwith (1997). "The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the Escherichia coli cytoplasm." Journal of Biological Chemistry **272**(25): 15661-15667.

57. Van Dat, N., F. Hatahet, K. E. H. Salo, E. Enlund, C. Zhang and L. W. Ruddock (2011). "Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E-coli." Microbial Cell Factories **10**.

58. Baneyx, F. (1999). "Recombinant protein expression in Escherichia coli." Current Opinion in Biotechnology **10**(5): 411-421.

59. Geertsma, E. R. and R. Dutzler (2011). "A Versatile and Efficient High-Throughput Cloning Tool for Structural Biology." Biochemistry **50**(15): 3272-3278.

60. Buchan, D.W., Ward, S.M., Lobley, A.E., Nugent, T.C., Bryson, K. and Jones, D.T. (2010) "Protein annotation and modelling servers at University College London". Nucl. Acids Res. 38 Suppl, W563-W568.

61. Jones DT. (1999) "Protein secondary structure prediction based on position-specific scoring matrices". J. Mol. Biol. 292: 195-202.

62. Adamczak R., Porollo A. and Meller J., (2004) "Accurate Prediction of Solvent Accessibility Using Neural Networks Based Regression", Proteins: Structure, Function and Bioinformatics, 56:753-67.

63. Adamczak R., Porollo A.and Meller J., (2005) "Combining Prediction of Secondary Structure and Solvent Accessibility in Proteins", Proteins: Structure, Function and Bioinformatics, 59:467-75.

64. Wagner M., Adamczak R., Porollo A. and Meller J., (2005) "Linear regression models for solvent accessibility prediction in proteins", Journal of Computational Biology, 12:355-69.

65. Porollo A., Adamczak R., Wagner M. and Meller J., (2003) "Maximum Feasibility Approach for Consensus Classifiers: Applications to Protein Structure Prediction", CIRAS (conference proceedings).

66. Cole C, Barber JD and Barton GJ. (2008) "The Jpred 3 secondary structure prediction server", Nucleic Acids Res,. 35: 197-201.