2013

DEPARTAMENTO DE CIÊNCIAS DA VIDA

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

A Chloroplastidial Atypical Aspartic Protease from A*rabidopsis thaliana*:

Optimization of Heterologous Expression, Purification and Biochemical Characterization

Daniela Patrícia Martins Dias Pedroso

2013

A Chloroplastidial Atypical Aspartic Protease from Arabidopsis thaliana:
Optimization of Heterologous Expression, Purification and Biochemical Characterization

Daniela P.M. Dias Pedroso

DEPARTAMENTO DE CIÊNCIAS DA VIDA

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

A Chloroplastidial Atypical Aspartic Protease from A*rabidopsis thaliana*:

Optimization of Heterologous Expression, Purification and Biochemical Characterization

Dissertação apresentada à Universidade de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Bioquímica, realizada sob a orientação científica do Professor Doutor Euclides Pires (Universidade de Coimbra) e da Doutora Isaura Simões (CNC/Biocant)

Daniela Patrícia Martins Dias Pedroso

2013

"E assim sou, fútil e sensível, capaz de impulsos violentos e absorventes, maus e bons, nobre e vis (…). Tudo em mim é a tendência para ser a seguir outra coisa; uma impaciência da alma consigo mesma, como uma criança inoportuna; um desassossego sempre crescente e igual. Tudo me interessa e nada me prende. Atendo a tudo sonhando sempre (…)."

Fernando Pessoa, **Livro do Desassossego**

Agradecimentos

Finalizada esta etapa, resta-me agradecer a todas as pessoas que, de uma forma ou outra contribuíram para que este caminho fosse possível.

Ao Professor Carlos Faro e ao Professor Euclides Pires por tornarem possível a realização deste trabalho nas condições ideais e pelo optimismo que sempre transmitem.

À Professora Isaura Simões, por ter aceitado orientar-me neste dois anos. Pelo apoio incondicional a cada passo. Por ter compartilhado comigo o seu conhecimento e por ter contribuído de forma tão rigorosa e crítica para a minha formação. Pela disponibilidade com que recebeu as minhas dificuldades e pela forma optimista com que as converteu em soluções. Pelo entusiasmo demonstrado a cada etapa conquistada. Agradeço pela liderança que a todos motiva para fazer cada vez mais e melhor.

A todos na Unidade de Biotecnologia Molecular, ao Doutor Pedro Castanheira, à Carla, à Marisa, ao André parceiro nesta luta com as atípicas, ao Rui que tudo sabe e tudo resolve, à Joana, à Ana Sofia, à Rita e ao Pedro. Pela ajuda no meu trabalho, pelo bom ambiente vivido e pelas inúmeras tardes de riso e muitas gomas.

Aos meus pais por todo o amor e dedicação ao longo deste percurso. Obrigada pelo apoio em cada etapa e por me motivarem a querer sempre mais, a ir mais além.

Ao meu irmão pelo amor que tem por mim, pela preocupação e amizade. O esforço e dedicação que colocas em tudo o que fazes são para mim um exemplo.

À minha avó Luz e ao meu padrinho Nuno que sempre me apoiaram e que partiram cedo demais para poder ver esta conquista. *Consegui!!!*

A dois antigos professores, Professos Filipe Menezes e Professor José Eduardo Ferraz. Por fazerem crescer em mim este bichinho inconformista, com vontade de pensar mais, sem barreiras e preconceitos. Por fazerem nascer em mim o amor à Academia, a Coimbra.

Às minhas "irmãs emprestadas" Cris, Teresa e Vera que provam todos os dias que grandes amizades não esmorecem com o tempo.

Deixo um abraço especial para um grupo de grandes amigos, que se foram juntando ao acaso à volta das diversas mesas do Tropical. Ao Paulo que muito me aturou este ano, obrigada pela amizade e paciência e por todas as conversas ao sabor de mais um café. Ao Ricardo por conseguir trazer luz aos meus dias mais negros, pela forma alegre com que vive a vida, contagiando-nos a todos. À Paulinha, companheira de chá e de grandes conversas, à "Suuusaaanaaaa" pela constante alegria, à Miqui, à Anna, sincera amiga, e ao Caramelo. Ao Alfa pelas geniais discussões sobre tudo; ao Beta pela verdadeira amizade e por fomentar o meu vício em café.

Finalmente, agradeço a todos os amigos que contribuíram para eu ter chegado até aqui e a todos os inúmeros colegas de Bioquímica com quem compartilhei este 5 anos em Coimbra.

Table of Contents

Abbreviations

A280nm - Absorbance at 280nm

ABA- Abcissic Acid

AP - Aspartic Protease

ASPG1-Aspartic Protease in Guard Cells 1

CDR1 - Constitutive Disease Resistance 1

CND41-Chloroplast Nucleoid DNA binding protein

DNA - Deoxyribonucleic Acid

DNP - 2,4-Dinitrophenyl

DTT - Dithiothreitol

E-64 - Trans-Epoxysucciny-L-leucyl-amido (4-guanidino) butane

EDTA - Ethylenediaminetetraacetic acid – Endoplasmic Reticulum

FPLC - Fast Protein Liquid Chromatography

GSH- Reduced Glutathione

GSSG-Oxidized Glutathione

HPLC-High Performance Liquid Chromatography

IMAC - Immobilized Metal Ion Affinity Chromatography

IPTG - Isopropyl β-D-1-thiogalactopyranoside

kDa - Kilodalton

LB – Luria Broth

LBAmp - Luria Broth with Ampicillin

LBAmp/Cam - Luria Broth with Ampicillin and Chloramphenicol

MCA - (7-methoxycoumarin-4-yl)acetyl

min – Minutes

mL-Milliliters

Mut - Mutant

OD600nm - Optical density at 600nm

PVDF - Polyvinylidene Fluoride

PCR- Polymerase Chain Reaction

PCS1- Promotion of Cell Survival 1

pI – Isoelectric Point

PSI- Plant-Specific Insert

PvNod41- Phaseolus vulgaris Nodulin 41

RP-HPLC – Reversed-Phase High-Performance Liquid Chromatography

ROS- Reactive Oxygen Species

SA- Salicylic Acid

SAR- Systemic Acquired Resistance

SDS - Sodium Dodecyl Sulfate

SDS-PAGE - Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

TB - Terrific Broth

TBS - Tris Buffered Saline

TBST - Tris Buffered Saline with Tween 20

TFA - Trifluoroacetic Acid

TN- Tris-NaCl

TNT- Tris-NaCl with 1% of Triton X-100

V -Volts

WT - Wild type

Abstract

Aspartic proteases are widely distributed among all the biological kingdoms. Plants have a higher number of aspartic proteases, suggesting that in these organisms aspartic proteases may play an important role in different biological processes.

Plant atypical aspartic proteases have very unique characteristics. They present a very different domain organization in its primary structure (loss of the PSI domain), a higher number of cysteine residues and a unique amino acid sequence around the first catalytic triad (hydrophobic-hydrophobic-DTG-Serine-Acid). These proteases were found in very unusual localizations and several authors claim that they have specialized functions and are involved in highly regulated processes.

NP_181826 (At2g42980) is a chloroplastidial atypical aspartic protease from *Arabidopsis thaliana*. This protease has a putative DNA-binding domain, suggesting that DNA may have a regulatory influence in its catalytic activity. Also, a bipartite presequence is present in NP_181826 primary sequence. These bipartite presequences were only found so far in organisms that experienced a secondary endosymbiosis and they were never described in superior plants, such as *Arabidopsis*. The chloroplastidial localization and its unusual sequence features suggest that this aspartic protease may have a critical role in the biological system.

The aim of this study was to establish and optimize an expression and purification protocol for production of NP_181826, using *E. coli* as the heterologous host. A putative mature form of NP_181826 (35toDTG) was cloned and expressed in this system.

Several approaches were used to optimize 35toDTG expression and purification. When expressed as soluble protein, 35toDTG was shown to co-purify with several other proteins and the protein yields were always very low. Expression as inclusion bodies followed by refolding by dialysis was the best strategy to obtain a higher yield of purified protein.

Recombinant 35toDTG showed proteolytic activity towards a typical aspartic protease substrate, at an optimum pH value of 3.5. This protease showed a very restricted specificity, being able to cleave oxidized insulin β-chain in only one site. Its complete

inhibition in the presence of Pepstatin A and the total loss of proteolytic activity observed for the active-site mutant, positions NP_181826 in the A1 family of aspartic proteases. The inhibition effect of DTT and GSH, with no effect in the presence of GSSG, and NADH and ATP activation effect in 35toDTG proteolytic activity suggest that the redox environment and the presence of nucleotide species can have a direct effect on the stability and/or catalytic activity of this atypical aspartic protease.

These experimental evidences strongly suggest that NP_181826 may have a very specific function in *Arabidopsis*'s chloroplast and that the redox environment of the chloroplast may have a direct influence on the protease activity.

Keywords

Plant aspartic proteases; Atypical aspartic proteases; Chloroplast; *Arabidopsis thaliana*

Resumo

As proteases aspárticas estão amplamente distribuídas por todos os reinos biológicos. O reino das plantas apresenta um número mais elevado de proteases aspárticas do que os restantes reinos biológicos, o que sugere que estas possam ter um papel importante nos seus processos biológicos.

As proteases aspárticas atípicas de plantas possuem características únicas. Apresentam uma organização muito diferente da sua estrutura primária (não possuem o domínio PSI), possuem um elevado conteúdo em cisteínas e uma sequência de aminoácidos única à volta da primeira tríade catalítica (hidrofóbico-hidrofóbico-DTG-serina-acídico). Estas proteases têm localizações celulares pouco usuais e vários autores afirmam que estas enzimas possuem uma função especializada, estando envolvidas em processos altamente regulados.

A NP_181826 (At2g42980) é uma protease aspártica atípica cloroplastidial de *Arabidopsis thaliana*. Esta protease possui um domínio de ligação a DNA, o que sugere que o DNA poderá ter um efeito de regulação da actividade da protease. Esta protease também apresenta uma presequência bipartida. Este tipo de presequência só foi descrito em organismos que sofreram um processo de endossimbiose duplo, sendo que nunca foi encontrada em plantas superiores como a *Arabidopsis thaliana*.

As suas características únicas e o facto de se encontrar no cloroplasto sugerem que esta protease aspártica possa ter um papel extremamente importante neste organelo.

O objectivo deste trabalho consistiu em estabelecer e optimizar protocolos de expressão e purificação para a produção da NP_181826, usando *E.coli* como sistema heterólogo de expressão. Neste estudo, uma putativa forma madura da NP_181826 (35toDTG) foi clonada e expressa neste sistema.

Várias estratégias foram utilizadas para optimizar a sua expressão e purificação. Quando expressa na forma solúvel, a forma recombinante 35toDTG apresentou tendência para ser co-purificada com outras proteínas e a quantidade de proteína obtida foi muito baixa. A expressão na forma de corpos de inclusão seguida de um

processo de *refolding* por diálise resultou na melhor estratégia para obter quantidades elevadas de proteína purificada.

A proteína recombinante 35toDTG apresentou actividade proteolítica contra o substrato típico das protease aspárticas, com um pH óptimo a 3.5. Esta protease aparenta ter uma especificidade muito restrita, sendo capaz de cortar a cadeia β oxidada da insulina em apenas um local. A sua completa inibição na presença de Pepstatina A e a perda total de actividade do seu mutante do local activo colocam esta protease na família A1 das proteases aspárticas. A inibição com DTT e GSH, não havendo qualquer inibição com GSSG, e o efeito de activação por parte do NADH e ADT levam a concluir que o ambiente redox e a presença de nucleótidos podem influenciar directamente a estabilidade e/ou actividade catalítica desta protease aspártica atípica.

Os resultados experimentais sugerem que a NP_181826 poderá ter uma função altamente específica no cloroplasto da *Arabidopsis* e que o ambiente redox do cloroplasto poderá influenciar directamente a sua actividade proteolítica.

1. Introduction

1.1 Plant Aspartic Proteases

Aspartic proteases are widely distributed among all the biological kingdoms. They can be found in vertebrate, plants, yeast, fungi, parasites, nematodes, bacteria and even in virus. Since aspartic proteases were described as a specific class of proteinases, they have been intensively characterized. It has been demonstrated that aspartic proteases play key roles in very different processes, such as protein processing, protein degradation or viral polyprotein processing (Dunn 2002; Simões & Faro 2004). Also, several aspartic proteases are known to be involved in various disease conditions, including malaria, AIDS, hypertension and Alzheimer's disease (Coates et al. 2008).

According to the MEROPS database, aspartic proteases can be divided into 16 families and 5 clans, according to their similarity in amino acid sequences and structural relationships, respectively (Rawlings et al. 2012). Looking closely to their distribution among different biological kingdoms, it can be easily concluded that plants have a higher content of aspartic proteases, suggesting that these enzymes have an important role in plants' biological functions.

Unlike mammalian and microbial proteases, plant aspartic proteases are not as well characterized and their biological functions are still hypothetic, in most of the cases. They were implicated in several important biological functions in plants' life (plant senescence, responses to abiotic or biotic stress, processes of programmed cell death and plant reproduction), but so far, for most of them, it hasn't been possible to define the specific role they play in these physiological processes (Xia et al. 2004; Olivares et al. 2011; Ge et al. 2005; Phan et al. 2011; Chen et al. 2008; Kato et al. 2005; Athauda et al. 2004; Yao et al. 2012; Gao et al. 2011; Simões & Faro 2004).

Plant aspartic proteases have been grouped in 3 different clans: AA (families A1, A2, A3, A11, A28, A33), AC (A8) and AD (A22, A24). Most plant aspartic proteases can be found in A1 family, the pepsin-like family (Rawlings et al. 2012), and are characterized by having a common catalytic apparatus and a conserved DTG/DSG sequence (Simões & Faro 2004). In general, and like other A1 aspartic proteases, they are active at acidic pH, are inhibited by pepstatin A - a natural hexapeptide from *Streptomyces* (Faro & Gal

2005) – and display specificity for hydrophobic residues (Phe, Val, Ile, Leu) at P1 and P1' positions (Vairo Cavalli et al. 2013).

The catalytic mechanism of A1 plant aspartic proteases is based on the interaction of two aspartic acid residues present in the catalytic site and on a nucleophilic attack of a water molecule, polarized by the aspartic acid residue of the second triad (Coates et al. 2008). The water nucleophilic attack of the scissile carbonyl bond results in a tetrahedral intermediate, stabilized by hydrogen bonds to the negatively charged carboxyl from aspartic acid residue of the first triad (Fig.1). Protein cleavage is accompanied by a proton transfer to the leaving group (Coates et al. 2008; Dunn 2002).

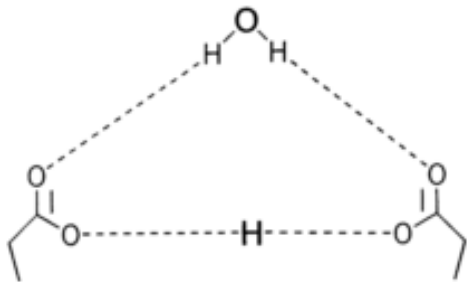

Figure1. Tetrahedral intermediate formed in aspartic proteases catalytic mechanism (Dunn 2002)

1.2 Atypical Aspartic Proteases

1.2.1 Sequence and structural features of plant aspartic proteases

Typically, plant aspartic proteases are synthesized as single-chain preproenzymes with an N-terminal signal sequence, involved in translocation to the ER, and a prosegment. This last domain is involved in inactivation, correct folding, stability or intracellular sorting of proenzymes (Simões & Faro 2004). Between the protease's N-terminal domain and C-terminal domain, it can also be found a 100 amino acid domain, the plant-specific insert (PSI) (Simões & Faro 2004) (Fig.2).

Figure 2. Schematic representation of the primary structure organization of typical plant aspartic proteases. Distinguishing features of plant aspartic proteases are highlighted (**SP**- signal peptide; **Pro**-prosegment; **DTG**- first catalytic triad; **PSI**- Plant-specific insert **DSG**- second catalytic triad) (adapted from (Faro & Gal 2005)**).**

This domain, identified exclusively in plant aspartic protease zymogens, is highly similar to saposins and saposin-like proteins (SAPLIPs). The PSI domain has six conserved cysteines, several hydrophobic residues, including a highly conserved tyrosine related to the definition of the catalytic pocket, and a consensus glycosylation site (Simões & Faro 2004; Faro & Gal 2005). PSI can fold as an independent domain. Structural characterization of phophytepsin's PSI and other proteins from the SAPLIP family shows that they share a compact globular structure, with five α-helices, linked by three disulfide bonds (Simões & Faro 2004; Bryksa et al. 2011). In fact, PSI can't be defined as a true saposin. Although it presents all the characteristic features of SAPLIPs, the PSI domain has a swap of the N-terminal and C-terminal portions of the saposin-like domain, making it a swaposin instead. This swap event is a clear example of a very rare evolutionary process in which one of the homologous sequences has the exactly same conserved domains as the other homologous sequences, but they are present in a different sequential order (Simões & Faro 2004; Ponting & Russell 1995). In the PSI domain this corresponds to a permutation of amino-terminal and carboxyl-terminal saposin fragments, being the C-terminal portion linked to the N-terminal portion (Egas et al. 2000).

The function of the PSI domain hasn't been completely clarified yet. Several authors have shown that this domain, as described for other saposins, can interact with lipid bilayers (Egas et al. 2000; Bryksa et al. 2011). Also, it was suggested that PSI can be involved with the sorting of zymogens during Golgi-mediated intracellular transport to the vacuoles, having an early strong influence in the interaction of the protein precursors with the ER and then allowing its interaction with membranes and membrane-bond receptor proteins in the Golgi (Törmäkangas et al. 2001). The role of vacuolar sorting signal was also described for the mammalian saposin C (Simões & Faro

2004). Other results indicate that PSI's ability to interact with membranes make it a potent inducer of vesicle leakage, suggesting that it can be part of a defense mechanism against pathogens or can even be an effector of cell death (Simões & Faro 2004). Munõz et al recently showed that *Solanum tuberosum* aspartic protease 1 PSI (*St*Asp-PSI) is able to kill microbial pathogens thanks to direct interaction with the microbial cell wall, leading to permeability and lysis (Muñoz et al. 2010).

In order to become mature, plant aspartic protease zymogens need to undergo several cleavage processes. In the end, single-chain or two-chain mature forms are produced. After translocation to the ER lumen, plant aspartic preproproteases suffer a cleavage of the signal peptide, followed by the cleavage of the prosegment and total or partial cleavage of the PSI domain, turning into a mature protease (Simões & Faro 2004). Still, the processing mechanism is not necessarily always the same and does not necessarily follows the same order (Simões & Faro 2004).

Conversion of the zymogen into an active aspartic protease, due to cleavage of the prosegment, can be an autocatalytic process, requiring a drop in pH, or it may need the presence of other molecules, such as other proteases or exopeptidases to promote the activation mechanism (Simões & Faro 2004).

Sequencing of *Arabidopsis thaliana* genome has given a lot of new information about aspartic proteases primary structure. In fact, as Faro and Gal described using an *in silico* analyses of Arabidopsis genome, most of the putative aspartic proteases differ from the basic structure described above (Faro & Gal 2005).

Based on these results, aspartic proteases can be currently divided in three different groups: typical, nucellin-like and atypical aspartic proteases. The members of typical aspartic proteases present the basic primary structure previously described; thus their sequences code for a PSI domain, a signal peptide and a prosegment (Faro & Gal 2005). Most of the aspartic proteases studied and characterized so far belong to this group and, until recently, this domain organization was the definition itself of plant aspartic proteases.

The knowledge of the existence of atypical aspartic proteases and nucellin-like aspartic proteases rewrite the concept of plant aspartic proteases. In fact, atypical and nucellin-like aspartic proteases don't have a PSI domain (Fig3).

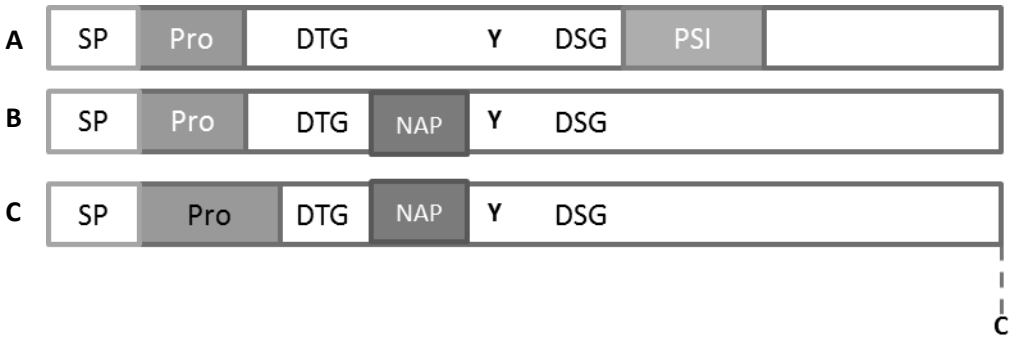

Figure3. Distinguishing aspects of Typical [A], Nucellin-like [B] and Atypical [C] aspartic proteases: Active center sequences Distinguishing features of plants' aspartic proteases are highlighted (**SP**- signal peptide; **Pro**-prosegment; **DTG**- first catalytic triad; **NAP**- Nepenthesin-type aspartic protease specific insertion; **Y-** conserved Tyr ; **DSG**- second catalytic triad; **C**- Cysteine) (Faro & Gal 2005).

Nucellin-like aspartic proteases also show significant deletions in the prosegment region (Faro & Gal 2005).

With the exception of 41-kD Chloroplast Nucleoid DNA binding protein (CND41) and Promotion of Cell Survival 1 (PCS1), all the atypical Aspartic proteases characterized so far follow almost a typical structural architecture with some variations: a signal peptide, follow by a prosegment and a mature protein coding domain.

CND41 (chloroplast nucleoid DNA binding protein) has a putative long transit peptide (120 amino acids), involved in the protease sorting to the chloroplast (Nakano et al. 1997). Also, PCS1 (Promotion of Cell survival 1) seems to lack a classic prosegment. Instead, after a signal sequence, it has a serine-rich sequence, whose function hasn't been described so far (Faro & Gal 2005).

Also, and as Faro and Gal's results previously suggested, in atypical aspartic proteases the length of the prosegment domain can be very different. Constitutive disease resistance 1 (CDR1,) has a 46 amino acid putative prosegment and it is active without its irreversible removal (Simões et al. 2007). On the other hand, nodulin 41 (PvNod41) has a 50 amino acid prosegment and nepenthesins I and II present a 56 and 55 amino

acid prosegments, respectively (Simões et al. 2007; Olivares et al. 2011; Athauda et al. 2004).

Two three-dimensional structures of plant aspartic proteases have been determined so far, both belonging to the typical group: cardosin A mature form (Frazao 1999) and prophytepsin, with its prosegment and PSI domain (Kervinen et al. 1999). Their structure is similar to each other and to what was determined for other aspartic proteases. It consists of a bilobal structure, composed by two polypeptide chains, with essentially β-strands and a few α-helixes. The active site is located between the two lobal domains, each one having a catalytic triad. Also, there is a flexible structure, the flap, over the catalytic cleft that interferes directly with the protease specificity, allowing the entrance of the protease substrate (Simões & Faro 2004; Vairo Cavalli et al. 2013). Within this flap there is a conserved Tyr (Tyr75, pepsin numbering), defining the catalytic pocket that interacts with substrate side-chain residues. So far, no atypical aspartic protease structure was determined and so all considerations about their structure and implications related to it are made considering typical aspartic protease structures.

1.2.2 Specific sequence around the first catalytic triad and cysteine content

Along with all the different domain organization features, atypical, nucellin-like and typical aspartic proteases also have characteristic sequences around the first catalytic triad (Table1) (Faro & Gal 2005).

Table 1. **Typical, nucellin-like and atypical aspartic proteases: different characteristic sequences around the first catalytic triad**

	First catalytic triad sequence
Typical AP	Hydrophobic-Hydrophobic-**DTG**-Serine-Serine
Nucellin-like	Acidic-Hydrophobic-**DTG**-Serine-Acidic
Atypical	Hydrophobic-Hydrophobic-**DTG**-Serine-Acidic

The first catalytic triad from typical aspartic proteases is in a Hydrophobic-Hydrophobic-**DTG**-Serine-Serine sequence. Nucellin-like aspartic proteases, on the

other hand, present a very different sequence around the catalytic sequence: Acidic-Hydrophobic-**DTG**-Serine-Acidic. The presence of these two acidic residues around the **DTG** sequence may have a great influence in the catalytic environment and, consequently, in their specificity.

In atypical aspartic proteases the first catalytic triad is in a Hydrophobic-Hydrophobic-**DTG**-Serine-Acidic sequence (Faro & Gal 2005). This last acidic residue can also be important in the catalytic mechanism of these proteases, interfering with its substrate and/or cleavage specificity.

The cysteine content and distribution also differ in these three aspartic protease groups. In typical aspartic proteases, the PSI domain contains a cluster of six cysteines, creating three disulfide bridges, holding the PSI α-helices structure (Simões & Faro 2004; Bryksa et al. 2011) (see Fig3).

Unlike typical aspartic proteases, nucellin-like and atypical proteases share a very specific distribution of cysteine residues(Faro & Gal 2005). Although not displaying a PSI domain, there is a cluster of cysteines between the first catalytic triad and the conserved Tyr residue located in the flap region. This domain, defined as nepenthesin-type aspartic protease (NAP)-specific insertion, was first described in nepenthesins I and II. So far, its function has not been clarified (Simões & Faro 2004; Athauda et al. 2004). Also, these two groups of aspartic proteases have three cysteine residues downstream the second catalytic triad (Faro & Gal 2005). Atypical aspartic proteases also show a characteristic C-terminal cysteine residue at the extreme C-terminus of the protease domain (Faro & Gal 2005).

In addition to the uncommon distribution of cysteine residues, atypical and nucellin-like aspartic proteases also show an unusual high cysteine content. CDR1, an atypical aspartic protease, has twelve cysteines in its mature form. Results show that these cysteines may contribute to the stabilization of the homodimeric form of CDR1, via intermolecular disulfide bonds, which is critical for CDR1's catalytic activity (Simões et al. 2007). The authors also suggest that, *in vivo*, the formation of these disulfide bridges may depend on a redox-dependent mechanism, thereby regulating the monomeric/homodimeric balance of CDR1 and, consequently, its catalytic activity

(Simões et al. 2007). Similar to CDR1, it was also proposed that a change in redox environment during seed development may regulate the catalytic activity of FeAPL1, an atypical aspartic protease with twelve cysteine residues from *Fagopyrum esculentum* Moench (Milisavljevic et al. 2007; Milisavljevic et al. 2011). In nepenthesin I and II it can also be found a high cysteine content (twelve cysteines), being four of them in the NAP-specific domain (Athauda et al. 2004). The high cysteine content may be linked to the remarkable stability of both enzymes in extreme environments (Athauda et al. 2004).

1.2.3 Cleavage specificity and inhibition profile

In general, aspartic proteases show preferential specificity for cleavage between hydrophobic residues (e.g. Phe, Val, Ile, Leu) on either side of the scissile bond, at P1 and P1' positions, being Phe-Phe one of the most common cleavage sites (Vairo Cavalli et al. 2013). Even so, the cleavage preference of each protein is ultimately related to its natural substrates and, therefore, to its biological role.

Interestingly, atypical aspartic proteases show a very different range of cleavage specificity. CDR1 has extremely restricted cleavage specificity. It cleaves Leu-Tyr bond of oxidized insulin β-chain, at pH 6.0, being unable to cleave several model substrates, such as a typical aspartic protease substrate, BACE and renin substrates and other protein substrates, like BSA, hemoglobin and casein. It was necessary to design a specific fluorogenic substrate to characterize the proteolytic activity of CDR1. This substrate was cleaved at the Phe-Val bond, at pH 6.5. This is an unusual optimal pH for aspartic proteases proteolytic activity, but is in agreement with the unusual extracellular localization of the enzyme (Simões et al. 2007). Requirement of a Glu residue in the S4 pocket of CDR1 active site was suggested as the reason for the protease narrow substrate specificity (Simões et al. 2007).

Regarding PCS1, assays have shown that this protease has a broader specificity being able to cleave the oxidized insulin β-chain in six different sites, at pH 5.8. However, PCS1 displays a preference for aromatic and aliphatic amino acids in P1' position (Silva 2008).

Nepenthesin I shows a broader specificity as well, being able to cleave the oxidized insulin β-chain in more than five different places, at pH 3.0, showing a preference for the Leu residue in the P1 position (Athauda et al. 2004). Also, Leu-Cya bond (Cya - Cysteic acid) is one of the preferred bonds cleaved by nepenthesin I; so far it wasn't reported the cleavage of this bond by other aspartic proteases (Athauda et al. 2004).

Regarding aspartic protease in Guard Cells 1 (ASPG1), S5, PvNOd41, Dwarf Arabidopsis chloroplastidial aspartic protease (NANA) and CND41, few studies were done related to their cleavage specificity. Only a few assays were done to characterize PvNod41 proteolytic activity. PVNod41 presents a preferential binding to denatured proteins, such as BSA, lysozyme and α_2-macroglobulin, though it only shows a limited peptide activity against casein, at pH4.5 (Olivares et al. 2011). Similar to PvNod41, CND41 also show the ability to cleave casein but at pH 2.5, an unusually low pH that does not occur in normal physiological conditions in chloroplast and that was linked to the biological role of CND41 during leaves senescence (Murakami et al. 2000). Results suggest that CND41 may have a broad specificity, being able to act on different denatured chloroplast proteins, as well (Kato et al. 2004). ASPG1 and NANA also showed proteolytic activity with casein at an unusually high pH value (pH6.0), like CDR1, (Yao et al. 2012).

As it was said previously, plant aspartic proteases experience inhibition in the presence of Pepstatin A, a natural hexapeptide from *Streptomyces* (Faro & Gal 2005; Simões & Faro 2004). In fact, this inhibitor has been widely used to identify and purify aspartic proteases (Athauda et al. 2004). However, atypical aspartic proteases have very different sensitivities to pepstatin. ASPG1 and nepenthesin I and II suffer a strong inhibition by Pepstatin A (Athauda et al. 2004; Yao et al. 2012), whereas CDR1 and PvNod41 are only partially inhibited (Simões et al. 2007; Olivares et al. 2011). Regarding PCS1 and CND41, almost no inhibition was observed in the presence of pepstatin (Silva 2008; Murakami et al. 2000).

Several other inhibitors were tested towards atypical aspartic proteases. It was shown that Pefabloc, a classic serine protease inhibitor, has a partial inhibitory effect on CDR1. So far, inhibition of a plant aspartic protease with Pefabloc was never described. CDR1 also suffered inhibition with EDTA, suggesting the requirement of a metal ion to

proteolytic activity. Also some nucleotides and redox agents (oxidized and reduced glutathiones) were shown to inhibit CDR1. The negative effect of redox agents on CDR1 activity has been correlated with the stability of its homodimeric structure, which appears to be crucial to enzymatic activity (Simões et al. 2007). PCS1 showed also inhibition in the presence of EDTA and zinc, suggesting the necessity of a metal ion (Zinc ion) to PCS1's activity (Silva 2008). Also, a strong inhibition by reduced glutathione and complete inhibition in the presence of oxidized glutathione supports the hypothesis that cysteines may play a crucial role in PCS1 stability and activity (Silva 2008). The chloroplastidial CND41 is strongly inhibited by NADPH, ATP and other nucleoside triphosphates as well as Fe^{3+}. Since CND41 is involved in leaf senescence processes in chloroplasts, it is likely that photosynthetic products may regulate the activity of the protease (Murakami et al. 2000). Also DNA has an inhibitory effect on CND41 at pH 7.5 but doesn't have any effect at pH 2.5. Considering that CND41 has a DNA-binding domain, these results suggest that modification of a DNA-binding domain or of chloroplastidial DNA at stress conditions (pH 2.5) may regulate CND41 activity (Kato et al. 2004).

1.2.4 Unusual distribution and localization

Aspartic proteases are widely distributed among plants and they have been found in several plant tissues (Simões & Faro 2004).

Most of typical aspartic proteases are intracellular and they accumulate essentially inside vacuoles. Some typical aspartic proteases are also found in the extracellular medium (Simões & Faro 2004; Faro & Gal 2005). Cardosin B, a typical aspartic protease from *Cynara cardunculus* L., is found in the extracellular medium of the floral transmitting tissue (Simões & Faro 2004).

In addition to their unique structural features and different specificity and sensitivity to inhibitors, atypical aspartic proteases have been found in several unusual cellular localizations.

The predicted localizations of atypical aspartic proteases of *Arabidopsis thaliana*, described by Faro and Gal, are unexpected. Also, Chen and coworkers identified several atypical aspartic proteases from *Oryza sativa* L. with unusual localizations.

Some atypical aspartic proteases were predicted to be found in chloroplasts or have membrane localization. Others were predicted to be extracellular and to be located at the mitochondria, the endoplasmic reticulum membrane and cytoplasm (Faro & Gal 2005; Chen et al. 2009).

Nepenthesins I and II, examples of extracellular atypical aspartic proteases, can be found in pitcher fluid from *Nepenthes distillatoria*, a carnivorous plant (Athauda et al. 2004). CND41, from *Nicotiana tabacum* leaves, is the only DNA-binding aspartic protease described so far in chloroplasts (Nakano et al. 1997) Also, *Arabidopsis* NANA is found in the plant chloroplast (Paparelli et al. 2012). *Arabidopsis* PSC1, involved in programmed cell death processes, PvNod41, linked to Disease resistance in *Phaseolus vulgaris* L., and *Arabidopsis* ASPG1, related to resistance do drought stress, are retained in the endoplasmic reticulum (Ge et al. 2005; Olivares et al. 2011; Yao et al. 2012). *Arabidopsis* CDR1, also associated with disease resistance, is present in the apoplastic fluid (Xia et al. 2004) while UNDEAD protease, one of the direct targets of MYB80 and related to the pollen development and programmed cell-death, has a mitochondrial targeting signal (Phan et al. 2011).

1.2.5 Specialized function in highly regulated pathways

Atypical aspartic proteases unusual features and unexpected localizations suggest a specialized function for these aspartic proteases, unlike the unspecialized housekeeping role assigned to typical proteases (Simões et al. 2007; Chen et al. 2008) .

So far, only a few atypical aspartic proteases have been well described and studied. They seemed to be involved in highly regulated processes and evidences suggest that theses proteases may have a highly specialized function (Table 2).

Table2. Plant atypical aspartic proteases: localization and function. The atypical aspartic proteases present unusual localizations and are involved in highly regulated functions.

Protein	Localization	Function	Determination
CDR1 ***Arabidopsis thaliana***	Apoplastic fluid	**Disease Resistance**	Gain-of-function mutants[1]
PvNod41 ***Phaseolus Vulgaris* L.**	Endoplasmic reticulum		Isolation of proteins able to interact with a synthetic peptide derived from Nodulin 30[2]
PCS1 ***Arabidopsis thaliana***	Endoplasmic reticulum	**Programmed cell death**	Knockout line with an insertion in PCS1 coding gene[3]
UNDEAD ***Arabidopsis thaliana***	Mitochondria		Microarrays analysis of gene expression in MYB80 knockout mutants and in mutants whose fertility was restored[4]
S5 ***Oryza sativa* L.**	Cell wall	**Plant Reproduction**	Genetic analysis[5]
Nepenthesin (I,II) ***Nepenthes distillatoria***	Extracellular	**Protein Recycling**	Purification of carnivorous aspartic proteases in the pitcher fluid[6]
CND41 ***Nicotiana tabacum***	Chloroplast	**Leaf senescence**	Utilization of antisense CND41 tobacco lines (R28)[7]
At1g66180 ***Arabidopsis thaliana***	Chloroplast or Secretory Pathway	**Resistance to oxidative stress**	Microarray analysis to identify genes whose transcript levels respond to changes in leaf Ascorbic Acid content[8]
ASPG1 ***Arabidopsis thaliana***	Endoplasmic reticulum	**Resistance to drought stress**	Gain-of-function mutants, which ectopically overexpressed ASPG1[9]
NANA ***Arabidopsis thaliana***	Chloroplast	***Chloroplast Homeostasis***	Searching for mutants altered in their sugar-sensing abilities by screening T-DNA tagged lines from *Arabidopsis thaliana*[10]

[1](Xia et al. 2004), [2](Olivares et al. 2011), [3](Ge et al. 2005), [4] (Phan et al. 2011), [5] (Chen et al. 2008), [6] (Athauda et al. 2004), [7] (Kato et al. 2005), [8] (Gao et al. 2011), [9] (Yao et al. 2012), [10] (Paparelli et al. 2012)

A) Disease Resistance

Plant disease resistance processes involve complex mechanisms of recognition and defense responses against pathogens. The activation of defense responses leads to an

accumulation of reactive oxygen intermediates, nitric oxide and salicylic-acid (SA), potentiating defense mechanisms (Xia et al. 2004). SA is synthetized in response to several pathogens and is crucial to establish local and systemic-acquired resistance (SAR) (Loake & Grant 2007).

Arabidopsis CDR1 was associated with disease resistance against pathogenic attacks (Xia et al. 2004). CDR1 involvement in disease resistance is likely due to proteolytic generation of a peptide elicitor, which might be involved in the activation of SA-dependent local and systemic defense responses (Y. Xia et al. 2004). Its proposed role in formation of a peptide elicitor suggests that CDR1 may be involved in highly specific events requiring sequence specificity and tight activity regulation (Simões et al. 2007; Xia et al. 2004). So far it wasn't possible to clarify the process of generation of the elicitor, the elicitor itself and how it will activate the disease resistance mechanism.

Chatoo and coworkers proved the existence of a homologous of this protein in rice (*Oryza sativa* CDR1) with similar features as CDR1 and also involved in disease resistance processes (Prasad et al. 2009). Results suggested that OsCDR1 doesn't have an antifungal activity on its own. However, OsCDR1 induces Pathogenic-related (*PR*) gene expression (*PR*2), leading to a defense response mechanism (Prasad et al. 2010).

In the case of PvNod41, this protease is expressed in nitrogen-fixing root nodules during the symbiosis of *Phaseolus vulgaris* and rhizobia (Olivares et al. 2011). Sequence analysis of PvNod41 and CDR1 revealed that they are closely related, though its distribution within the cell resembles that of PCS1, the programed cell death atypical protease (Olivares et al. 2011). Although its function is still unknown, PvNod41 can be found exclusively in uninfected cells. This suggests that this protease might induce a defense response in uninfected cells in order to maintain its integrity and it can be speculated, due to its proximity, that PvNod41 may have a mechanism analogous to that of CDR1 (Olivares et al. 2011).

B) Programmed cell death

In plants, cell death is associated with normal seed development, gametogenesis and other developmental processes (Ge et al. 2005). Programmed cell death (PCD)

processes are a way to supply the young zygotic embryo and the young endosperm with nutrients (Bi et al. 2005; Milisavljevic et al. 2008; Milisavljevic et al. 2011).

PCS1 is an atypical aspartic protease with an important role as an anti-cell death component. It is relevant in determining the fate of cells in embryonic development and reproduction processes in *Arabidopsis* (Ge et al. 2005; Xu et al. 2010). Loss of PCS1 function by mutagenesis causes male and female gametophyte degeneration and excessive mortality in developing embryos. Studies have shown the presence of PCS1 in the endoplasmic reticulum (Ge et al. 2005). It can be speculated that PCS1 may act as an anti-cell death protease by processing and activating a survival factor polypeptide or by processing and inactivating a pro-PCD component in endoplasmic reticulum (Ge et al. 2005).

The *Arabidopsis thaliana* MYB80 is expressed in the tapetum and microspores and it is essential for tapetal and pollen development (Phan et al. 2011). Recently an aspartic protease (UNDEAD) was identified like one of the direct targets of MYB80. This aspartic protease belongs to A1 family and can be classified as an atypical protease. It is proposed that UNDEAD could hydrolyze an apoptosis-inducing protein in the mitochondria that participates in PCD. Therefore MYB80 could induce transcription of UNDEAD gene, preventing PCD. The timing of tapetal PCD is critical for pollen development, and any delay or inhibition results in male sterility (Phan et al. 2011).

C) Plant reproduction

The S5 triallelic system, has a profound relevance in the evolution and artificial breeding of cultivated rice, being responsible for compatibility of rice subspecies (Chen et al. 2008). In this triallelic system, S5i and S5j present in *indica* and *japonica* rice subspecies, respectively, encode atypical aspartic proteases, capable of forming homo- and heterodimers (Chen et al. 2008). Formation of dimers seems to be required to the atypical aspartic protease activity (Ji et al. 2012). The presence of S5i and S5j heterodimers leads to reduced fertility, blocking gene flow between *indica* and *japonica* rice subspecies (Ji et al. 2012). On the other hand, S5n, a wide-compatibility allele, is unable to dimerize because it has a deletion in the N-terminus (Chen et al.

2008). It is a non-functional allele and interactions between *indica* or *japonica* and wide-compatibility varieties produce fertile hybrids (Ji et al. 2012).

Ji and co-workers recently show that there is only a single amino acid difference between the aspartic protease in *indica* and *japonica* subspecies (Phe273 in S5i and S5n plants and Leu273 in S5j). In the presence of S5i and S5j heterodimers, this single difference may be responsible to interference in PCD processes and, consequently, in normal cell growth during gametogenesis and/or embryogenesis in rice ovary, leading to plant sterility (Ji et al. 2012).

D) Protein recycling and Leaf senescence

Senescence is a degenerative process, related to the reduction of photosynthetic capacity, chloroplasts degradation, loss of chlorophyll, degradation of proteins, lipids and nucleic acids and, in the end, cell death (Kato et al. 2005; Noodén 2013). Despite the degenerative nature of this process, it leads to recycling of nutrients and their translocation from senescence cells to reproductive organs and younger leaves, is crucial for plant development (Noodén 2013). Senescence is an extremely complex process, involving highly coordinated alterations in gene expression and structure (Kato et al. 2004).

CND41 is an atypical aspartic protease with an important role in senescence processes. This protease, found in tobacco cells, has the ability to bind DNA, thanks to a lysine-rich region, in the N-terminus (Kato et al. 2004; Nakano et al. 1997). Inhibition of CND41 activity by DNA, at pH7.5, suggests that degradation of chloroplastidial DNA or modification of the DNA-binding domain, at stress conditions, could be a trigger to activate CND41 (Kato et al. 2004). Reduction of sugar levels and nitrogen levels in senescence leaves and an increase in ribulose 1,5-biphosphate denaturation due to oxidative stress in senescent leaves can also be related to the activation of CND41 (Kato et al. 2005). Several reports show that the proteolytic regulation of ribulose 1,5-biphosphate by CND41 is the source of amino acids being transported to developing organs (Kato et al. 2004).

Nepenthesin I and II can be found in pitcher fluid from *Nepenthes distillatoria*, a carnivorous plant (Athauda et al. 2004). They have an important role in prey digestion and may be involved in protein recycling through degradation of proteins, present in preys, to obtain nitrogen, used by plants (Athauda et al. 2004).

E) Resistance to abiotic stress

Abiotic stress factors have a deep impact in living organisms (Brunet et al. 2009). These environmental conditions limit plant growth and productivity, changing biological processes, such as cell division and metabolic pathways. Plants have several defense strategies against abiotic stress (Saibo et al. 2009; Yao et al. 2012).

Ascorbic acid has several physiologic roles in plants, being known as a redox buffer and enzyme cofactor. It was advanced that it has important roles in stress responses and growth (Gao et al. 2011). Gao and coworkers have identified an aspartic protease whose levels increased after an increase of ascorbic acid pools and with light. It is proposed that this aspartic protease (At1g66180) belongs to the A1 family and it is strongly predicted that it may be classified as an atypical aspartic protease (Gao et al. 2011).

Under drought stress, *Arabidopsis thaliana* can adopt an Abcissic acid (ABA)-dependent pathway. In mediated ABA-pathways, elevation of ABA induces an increase of cytosolic calcium, activating anion channels in the plasma membrane of guard cells and leading to their closure. This elevation can also promote ROS production in the same cells and, consequently, the stomatal closure. ROS species can also be a second messenger to ABA pathways (Yao et al. 2012). ASPG1 is an atypical aspartic protease, found in guard cells endoplasmic reticulum. This protease seems to be important to plant survival under drought stress. Under these stresses conditions, elevation of ABA seems to induce ASPG1 expression, leading to stomatal closure and, consequently, avoiding water loss (Yao et al. 2012). This aspartic protease may be able to trigger ROS production in guard cell, contributing to ABA pathways signaling and resulting in guard cells closure. On the other hand, ASPG1 seems to have the ability to activate

antioxidases, like superoxide dismutase and catalase, preventing oxidative damage, due to excessive ROS accumulation, during drought conditions (Yao et al. 2012).

F) Chloroplast Homeostasis

Photosynthesis is a crucial process in plants life. Through photosynthesis, during the day, plants synthesize carbohydrates that are used in plant growing and morphogenesis. During the night, plant metabolic processes are ensured by the starch reserves that are accumulated during the day (Grennan 2006; Paparelli et al. 2012)
NANA is an atypical aspartic protease located in the chloroplast. Results suggest that its misexpression affects the photosynthetic carbon metabolism by affecting the endogenous sugar levels as wells as the starch metabolism (Paparelli et al. 2012). Also alterations of NANA expression seem to result in the down-regulation of some chloroplast-encoded genes. This may lead to a plastid retrograde signal, resulting in the down-regulation of some nuclear genes and, consequently, affecting chloroplast biological functions. This alterations can also be the reason for the dwarf phenotype, present in these plants (Paparelli et al. 2012). This results suggest that NANA may have a crucial role in chloroplast homeostasis (Paparelli et al. 2012).

1.3 NP_181826, a Chloroplastidial atypical aspartic protease from *Arabidopsis thaliana*

NP_181826 (At2g42980) is an atypical aspartic protease found in *Arabidopsis thaliana*'s chloroplasts (Ferreira, L. and Faro, R., unpublished results). It has a high content of cysteines (14 cysteines) and a unique amino acid sequence (hydrophobic-hydrophobic-DTG-Serine-Acidic) around the first catalytic triad. Also, analyses of NP_181826 primary sequence show a Lysine-rich region, which can probably be a DNA-binding domain (Fig.4).
This is not the first time that an atypical aspartic protease shows a putative DNA-binding domain. CND41 presents a DNA-binding region as well.

1 MSTKLSIFLL GLILFSVSPF SGDCRTLSGK HEHYSSSLNM FNSQDTMRFS SASSSTSNDC
61 GFSSKEHDPS KEHTRESVKP QSRIKQETKR TTHSVVDLQI QDLTRIKTLH ARFN***KSKKQK***
121 ***NEKVRKK***ITS DISLVGAPEV SPGKLIATLE SGMTLGSGEY FMDVLVGTPP KHFSLIL**DTG**
181 SDLNWLQCLP CYDCFHQNGM FYDPKTSASF KNITCNDPRC SLISSPDPPV QCESDNQSCP
241 YFYWYGDRSN TTGDFAVETF TVNLTTTEGG SSEYKVGNMM FGCGHWNRGL FSGASGLLGL
301 GRGPLSFSSQ LQSLYGHSFS YCLVDRNSNT NVSSKLIFGE DKDLLNHTNL NFTSFVNGKE
361 NSVETFYYIQ IKSILVGGKA LDIPEETWNI SSDGDGGTII **DSG**TTLSYFA EPAYEIIKNK
421 FAEKMKENYP IFRDFPVLDP CFNVSGIEEN NIHLPELGIA FVDGTVWNFP AENSFIWLSE
481 DLVCLAILGT PKSTFSIIGN YQQQNFHILY DTKRSRLGFT PTKCADI*

Figure 4. NP_181826 amino acid sequence. The signal peptide is underlined; the putative transit peptide is highlighted in gray; the cysteine residues are double underlined, the putative DNA-binding domain is in bold and italic and catalytic triads are in bold underlined.

Results suggest that DNA may have a regulatory role in CND41 activity, having an inhibitory effect under physiologic conditions. DNA degradation under stress conditions seems to result in CN41 activation. It is tempting to think that DNA may have a similar influence on NP_181826 proteolytic activity. Sequence alignment shows that NP-181826 is closely related to CND41 and CDR1 (Fig.5), sharing 32% and 28% similarity, respectively.

This protease appears to have a putative signal peptide and a putative chloroplastidial transit peptide, predicted using SignalP and ChloroP, respectively. These bipartite presequences are found in organisms that experienced a secondary endosymbiosis, through the uptake of an eukaryotic alga into an eukaryotic host cell and its reduction into an organelle, such as several organisms from algae (e.g. diatoms) (Gruber et al. 2007). In these organisms, the signal peptide and the transit peptide are individually involved in ER import and primary plastids import, respectively (Gruber et al. 2007). This bipartite sequence never was found the superior plants, such as *Arabidopsis thaliana*.

In superior plants, like *Arabidopsis thaliana*, protein target to plastids is based on a posttranslational import by the Toc/Tic mechanism and its interaction with the transit peptide (Gruber et al. 2007; Radhamony & Theg 2006). As the protein goes into the stroma, the transit peptide is then cleaved (Radhamony & Theg 2006).

Recently, studies of the chloroplast genome revealed several proteins that don't have a putative transit peptide.

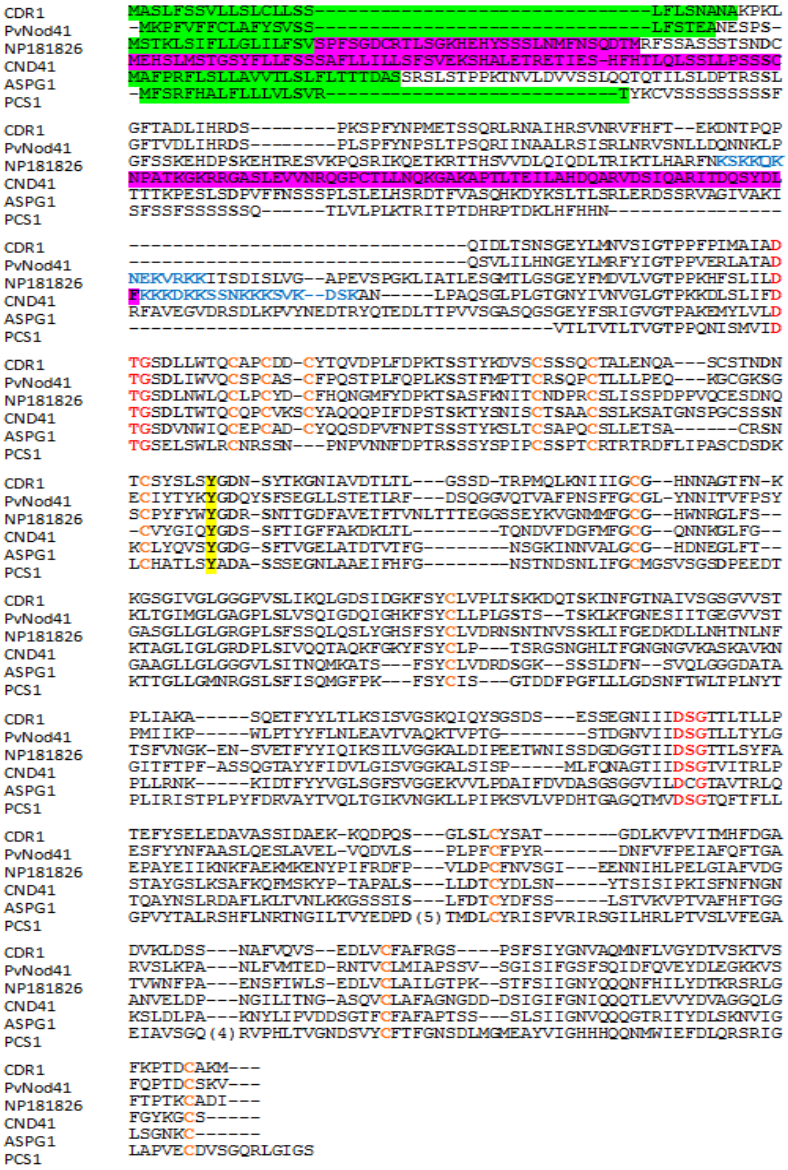

Figure 5. Multiple sequence alignment of NP_181826 and other atypical aspartic proteases. Alignment of NP_181826 with CDR1, (Accession number AAP72988), CND41 (BAA22813), PCS1 (NP_195839), PvNod41 (AEM05966) and ASPG1 (NP_188478) obtained with ClustalW2. The signal peptide is shown in green, the transit peptide in purple and the catalytic triads are in red. Also, the conserved flap tyrosine is shown in yellow and the cysteine residues in orange (Larkin et al. 2007; Goodstadt & Ponting 2001)

Chlapsin, a chloroplastidial aspartic protease from the green algae *Chlamydomonas reinhardtii*, and CAH1, an *Arabidopsis thaliana* α-carbonic anhydrase¸ don't have a transit peptide. Instead they present a signal peptide (Almeida et al. 2012; Radhamony & Theg 2006). Results suggest that these proteases may be involved in a newly identified ER-Golgi-plastid trafficking pathway (Almeida et al. 2012; Radhamony &

Theg 2006). Based on this, it is tempting to hypothesize that the bipartite sequence present in NP_181826 may also be responsible for its translocation to the chloroplast by an unusual pathway.

1.4 Objectives

Plant atypical aspartic proteases have very unique characteristics. They present a different domain organization in its primary structure, a higher amount of cysteine residues and a unique catalytic environment, due to their characteristic amino acid sequence around the first catalytic triad. These proteases were found in very unusual localizations and several authors claim that these aspartic proteases have specialized functions and are involved in highly regulated processes, crucial for the plant.

NP_181826 is a new chloroplastidial atypical aspartic protease from *Arabidopsis thaliana*. Analysis of its primary structure reveals very interesting aspects. This protease has a putative DNA-binding domain, suggesting that DNA can have a regulatory influence in NP_181826. The chloroplastidial localization of NP_181826 makes this hypothesis more attractive, since stressful conditions may have a direct effect on its catalytic activity, by leading to DNA degradation.
Also, this protease seems to have a bipartite presequence, only found so far in organisms that experienced a secondary endosymbiosis.

The chloroplastidial localization and its unusual sequence features suggest that this aspartic protease may have an important role in the biological system. Therefore, it is crucial to further characterize NP_181826 in order to understand its catalytic properties and to shed new light into its potential role in *Arabidopsis* chloroplasts.

For this purpose, the aim of this work is to optimize the heterologous expression of NP_181826 in *E. coli* cells and after that optimize and establish a purification protocol for the protease. These results will have great interest for future biochemical and structural studies. Also, a biochemical characterization of the enzyme will be performed in order to understand NP_181826 catalytic aspects, specificity and inhibition effect.

With this work, we expect to contribute to a better understanding of the properties of the eclectic group of atypical aspartic proteases.

2. Materials and Methods

2.1 Materials and Methods

Most of reagents used in this study were obtained from Sigma-Aldrich or Merck.

2.1 Cloning of different NP_181826 constructs

The coding sequence of the putative aspartic protease NP_181826 from *Arabidopsis thaliana* (accession number: NM_129859) was codon optimized for recombinant expression in *E. coli* and chemically synthetized by GenScript according to OptimusGene^{TM}Codon Optimization Analysis algorithm. Three different NP_181826 constructs in pET23a were then generated using the synthetic gene as the template.

For the first construct, corresponding to NP_181826 without the putative signal peptide and transit peptide (known as ΔSPTP), the DNA was PCR amplified with the forward primer 5' **CATATG**TTCTCCTCAGCATCGAGCTCT3' (with NdeI restriction sequence highlighted in bold) and the reverse primer 5'**CTCGAG**GATGTCGGCGCATTTGGTCG3' (with XhoI restriction sequence highlighted in bold).

For the second construct, corresponding to a putative mature form of NP_181826, starting 35 amino acids before the first catalytic triad DTG (named 35toDTG-His), DNA was amplified with the forward primer 5'**CATATG**GGCAAACTGATTGCTACGCTG3' (with an NdeI restriction sequence highlighted in bold) and the reverse primer 5'**CTCGAG**GATGTCGGCGCATTTGGTCG3' (with an XhoI restriction sequence highlighted in bold). Both set of primers allow these two amplicons to be subcloned into a pET23a vector (Novagen) in frame with a C-terminal His-tag.

The third construct, known as 35toDTG-Strep, corresponds to the putative protease mature form fused with a C-terminal Strep-tag. DNA was amplified using the following primers: The forward primer 5'**CATATG**GGCAAACTGATTGCTACGCTG3' (with a NdeI restriction sequence highlighted in bold) and the reverse primer 5'**CTCGAG**CTACTTTTCGAACTGCGGATGGCTCCA***ACCTCC***GATGTCGGCGCATTTGGTCG3'
(with a XhoI restriction sequence highlighted in bold, a Strep-tag coding sequence underlined and two spacer amino acids (two Gly) in bold and italic).

The amplified sequences were confirmed by 1% agarose gel electrophoresis and cloned in pCR™2.1 vector (TACloning®Kit- Invitrogen), following manufacturers' instructions. The ligation products were transformed in TOP10F' *E. coli* cells and positive colonies were chosen by blue-white selection and confirmed by DNA sequencing. The positive TA constructs and pET23a (Novagen) were then digested with NdeI and XhoI restriction enzymes. The inserts were purified by separation in 1% agarose gel electrophoresis, followed by DNA extraction with NZYGelpure kit (NZYTech). Digested and purified pET23a and inserts were incubated with DNA ligase. The ligation products were transformed in TOP10F' *E. coli* cells and the positive clones were selected by restriction analysis with NdeI and XhoI and confirmed by DNA sequencing.

The 35toDTG active-site mutant D178A (full sequence numbering), named 35toATG, was generated using the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene). The expression vector pET23a with 35toDTG-His insert was used as DNA template, and the forward primer 5'CATTTCTCACTGATCCTGGCTACCGGCTCGGACCTGAAT3' (mutation underlined) and reverse primer 5'ATTCAGGTCCGAGCCGGTAGCCAGGATCAGTGAGAAATG3' (mutation underlined) were used to obtain the active-site mutant, following manufacturers' instructions.

2.2 Expression Screening of ΔSPTP, 35toDTG-His and 35toDTG-Strep

For the expression screening of ΔSPTP and 35toDTG-His, the pET23a expression constructs were transformed in BL21Star (Invitrogen), C41 (Lucigen) and Origami (Novagen) chemically competent *E. coli* cells and in BL21Star (Invitrogen), C41 (Lucigen) and Origami (Novagen) co-transformed with a modified pLysS vector that includes Erv1p oxidase (sulfhydryl oxidase) and mature DsbC (disulfide bond isomerase) coding sequences under pBAD arabinose promoter and araC gene for regulation of the promoter (Hatahet et al. 2010; Nguyen et al. 2011). The transformed cells were inoculated in 5 mL Luria Broth (LB), with 100 μg/mL Ampicillin (LBAmp) or

with 100 µg/mL Ampicillin and 100µg/mL Chloramphenicol (LBAmp/Cam) when the cells were co-transformed with the modified pLysS vector, overnight at 37°C. The next day, the pre-inoculum was transferred to 15 mL LB or Terrific Broth (TB) medium with the respective antibiotics to an initial OD_{600nm} of 0.02. The cells were incubated at 37°C, with constant rotation. When the co-transformed cells reached an OD_{600nm} of 0.4, the expression of the folding factors was pre-induced by addition of 0.5% w/v Arabinose, followed 30 minutes later by the induction with 0.05 mM or 0.1 mM IPTG. The non-co-transformed cells were induced at an OD_{600nm} of 0.7 with 0.05 mM or 0.1 mM IPTG. After induction, cultures were incubated at 37°C for 3h or at 20°C for 3h or at 20° overnight, with constant rotation.

For the expression screening of 35toDTG-Strep, the pET23a expression construct was transformed in Origami (Novagen) co-transformed with the modified pLysS vector. The transformed cells were inoculated in 5 mL LBAmp/Cam (100 µg/mL Ampicillin and 100 µg/mL Chloramphenicol) overnight at 37°C. The next day, the pre-inoculum was transferred to 15 mL LBAmp/Cam medium to an initial OD_{600nm} of 0.02. The cells were incubated ate 37°C, with constant rotation. When the co-transformed cells showed a OD_{600nm} of 0.4, the expression of the folding factors was pre-induced by addition of 0.5% w/v Arabinose, followed 30 minutes later by the induction with 0.05 mM or 0.1 mM IPTG. After induction, cultures were incubated at 37°C for 3h or at 20° overnight, with constant rotation.

Total protein extraction was made using BugBuster Protein Extraction Reagent (Novagen). A volume of 1 mL of cell culture was harvested for use of protein extraction kit. The insoluble fraction was also ressuspended in 200 µL of PBS. Both soluble and insoluble fractions were analyzed by Western-blot.

2.3 Expression and purification of 35toDTG-His, 35toATG and 35toDTG-Strep as soluble protein

2.3.1 Expression of recombinant constructs

Co-transformed Origami cells with the modified pLysS vector and the 35toDTG-His pET23a construct were pre-inoculated in an Erlenmeyer flask with LBAmp/Cam medium (100 μg/mL Ampicillin and 100 μg/mL Chloramphenicol), at 37°C overnight, with constant rotation. The culture was then diluted (1:50) by Fernbach flasks containing 1 L LBAmp/Cam. The cells were incubated at 37°C with constant rotation to an OD_{600nm} of 0.4 and then pre-induced with 0.5% w/v Arabinose. After 30 minutes, protein expression was induced with 0.1mM IPTG. The culture was incubated at 37°C with constant rotation for 3h.

Expression of 35toDTG-His pET23a construct in high density flasks was made in a similar way. A co-transformed Origami pre-inoculum was incubated in Erlenmeyer flask with LBAmp/Cam medium, at 37°C with constant rotation overnight. The culture was again divided by Ultra Yield flasks (BioSilta), diluted (1:50) by each liter of LBAmp/Cam medium, and the cells were incubated at 37°C with constant rotation to an OD_{600nm} of 0.4 and then pre-induced with 0.5% w/v Arabinose. After 30 minutes, protein expression was induced with 0.1 mM IPTG and the cells were incubated at 30°C with constant rotation for 3h.

With a similar procedure, Origami cells co-transformed with the modified pLysS vector and the 35toDTG-Strep pET23a construct were pre-inoculated in an Erlenmeyer flask with LBAmp/Cam medium at 37°C, with constant rotation overnight. The pre-inoculum was diluted (1:50) in 1L LBAmp/Cam in Ultra Yield flasks (BioSilta). The cells were incubated as 37°C with constant rotation to an OD_{600nm} of 0.4 and then pre-induced with 0.5% w/v Arabinose. After 30 minutes, protein expression was induced with 0.1 mM IPTG. The culture was incubated at 20°C with constant rotation overnight.

For harvesting the cells, 1 L cultures were centrifuged at 6000 g for 20 min at 4°C. The pellet of 35toDTG expression was resuspended in 10 mL per 1 L of culture with 20 mM sodium phosphate buffer pH 7.5 with 10 mM Imidazole, 500 mM NaCl and then frozen at -20°C.

In the case of 35toDTG-Strep, the cell pellet was ressuspended in 10 mL per 1 L of culture with 100 mMTris/HCl buffer pH8.0 with 150 mM NaCl. The ressuspension was also frozen at -20°C.

2.3.2 Purification of 35toDTG-His by IMAC-Ni^{2+} and IMAC-Co^{2+}

Harvested cells expressing 35toDTG-His were thawed at room temperature and incubated with 0.01mg/mL DNAse and 20 mM $MgCl_2$ on ice until a fluid consistency was observed. The total protein extract was then processed using an EmulsiFlex C3 (Avestin) to cause cell lysis. The resulting lysates were centrifuged at 12000*g* at 4°C for 20 min. The soluble fraction was then filtered through 0.2 µm filters and applied to two His Trap HP 5 mL columns (Amersham Biosciences) charged with Ni^{2+}.

The two His Trap HP 5 mL columns were equilibrated and washed with 20 mM phosphate buffer pH 7.5, 10 mM Imidazole, 500 mM NaCl. For elution, 20 mM phosphate buffer pH 7.5, 500 mM Imidazole, 500 mM NaCl was prepared. The adequate percentage of elution buffer was mixed with the equilibration buffer to create the 50 mM, 100 mM and 500 mM Imidazole elution steps. The purification was carried out using a BioLogic DuoFlow FPLC system (Bio-Rad) at a constant flow of 5 mL/min. Fractions of 5 mL were collected through the elution steps.

Protein containing fractions were pooled and dialyzed using 12 kDa-cutoff dialysis membrane against 20 mM phosphate buffer pH 7.5 with 100 mM NaCl at 4°C, overnight.

The dialyzed pool was applied to His Trap HP 1 mL columns (Amersham Biosciences) charged with Co^{2+}. Reagent concentrations were then adjusted so that the pool sample would be in 20 mM phosphate buffer pH 7.5, 10 mM Imidazole, 500 mM NaCl. The purification step was carried out with a procedure similar to the previously described. The column was equilibrated and washed with 20 mM phosphate buffer pH 7.5, 10

mM Imidazole, 500 mM NaCl and the sample was eluted with 20 mM phosphate buffer pH 7.5, 500 mM Imidazole, 500 mM NaCl in three different steps: 50 mM, 100 mM 500 mM Imidazole elution steps. The purification was carried out using a BioLogic DuoFlow FPLC system (Bio-Rad) at a constant flow of 1 mL/min. Fractions of 1 mL were collected through the elution steps.

Protein containing fractions were dialyzed using 12 kDa-cutoff dialysis membrane against 20 mM Tris/HCl buffer pH 7.5 with 50 mM NaCl at 4°C overnight.

2.3.3 Purification of 35toDTG-His by IMAC-Ni^{2+} and Anionic Exchange Chromatography

Harvested cells expressing 35toDTG-His were thawed at room temperature and were processed as described before. The soluble fraction was filtered through 0.2 µm filters and applied to a His Trap HP 5 mL column (Amersham Biosciences) charged with Ni^{2+}.

The His Trap HP 5 mL column was equilibrated and washed with 20 mM phosphate buffer pH 7.5, 10 mM Imidazole, 500 mM NaCl and the sample was eluted with 20mM phosphate buffer pH 7.5, 500 mM Imidazole, 500 mM NaCl in three different steps: 50 mM, 100 mM 500 mM Imidazole elution steps. The purification was carried out using a BioLogic DuoFlow FPLC system (Bio-Rad) and, when the protein was expressed with Ultra Yield flasks, the purification was carried out using an AKTA FPLC system (Ge Healthcare). The protein was purified at a constant flow of 5 mL/min. Fractions of 5 mL were collected through the elution steps.

Protein containing fractions were pooled and dialyzed using a 12 kDa-cutoff dialysis membrane against 20 mM Tris/HCl buffer pH 7.5 with 100 mM NaCl at 4°C overnight.

The dialyzed pool was applied to an anion exchange Mono-Q 5/50 GL column (Amersham Biosciences). The column was equilibrated and washed with 20 mM Tris/HCl buffer pH7.5 and the protein was eluted with a linear gradient of NaCl (0 to 1 M of NaCl) in the same buffer. The purification was carried out using a BioLogic DuoFlow FPLC system (Bio-Rad) and, when the protein was expressed with Ultra Yield flasks, the purification was performed in an AKTA FPLC system (GE Healthcare). The protein was purified at a constant flow of 0.75 mL/min. Fractions of 1 mL were collected through the elution steps.

2.3.4 Purification of 35toDTG-Strep by a Strep-Tag affinity chromatography

Harvested cells expressing 35toDTG-Strep were thawed at room temperature and were processed as described before. The soluble fraction was filtered through 0.2µm filters and applied to StrepTrap HP 1mL column (GE Life Sciences).

The StrepTrap HP 1mL column was equilibrated and washed with 100 mMTris/HCl buffer pH 8.0 with 150 mM NaCl and the sample was eluted 100mMTris/HCl buffer pH 8.0 with 150 mM NaCl and 2.5 mM desthiobiotin in one elution step. The purification was carried out using an AKTA FPLC system (Ge Healthcare) at a constant flow of 0.75mL/min. Fractions of 1 mL were collected through the elution step.

2.4 Expression, refolding and purification of 35toDTG-His and 35toATG from inclusion bodies

2.4.1 Expression as inclusion bodies

BL21 Star cells transformed with the constructs 35toDTG-HispET23a and 35toATG active-site mutant constructs were pre-inoculated in an Erlenmeyer flask with ZBAmp medium (10 g/L N-Z amine and 5 g/L NaCl with 100 µg/mL ampicillin) at 30°C with constant rotation overnight. The culture was then diluted (1:50) in 1 L LBAmp in Fernbach flasks. The cells were incubated at 37°C with constant rotation to an OD_{600nm} of 0.7 and then induced with 0.5 mM IPTG. The culture was incubated at 37°C with constant rotation for 3h.

For harvesting the cells, the cultures were centrifuged at 6000 g for 20 min at 4°C.

The culture pellet was resuspended in 50 mMTris/HCl buffer pH 7.5, 50mM NaCl (TN buffer) and then was frozen at -20°C.

2.4.2 Refolding of recombinant 35toDTG-His and 35toATG

After thawing, the harvested cells expressing 35toDTG-His and 35toATG mutant were incubated with 0.01 mg/mL DNase and 20 mM $MgCl_2$ on ice until a fluid consistency was obtained. The total protein extract was then processed using an EmulsiFlex C3 (Avestin) to promote cell lysis. The cell lysate was diluted in TN buffer and washed overnight at 4°C with agitation. Then, the sample was centrifuged at 6000 g for 20 minutes at 4°C and washed with TN buffer containing 0.1 % (v/v) Triton X-100 (TNT), overnight at 4°C with agitation. After centrifugation at 6000 g for 20 minutes ate 4°C, the inclusion bodies were dissolved in 8 M Urea Buffer (8 M Urea, 0.1 M Tris, 1 mM Glycine buffer pH10.5, 1 mM EDTA) with 7 mL/L β-mercaptoethanol. The denatured proteins were then refolded by two different strategies.

2.4.2.1 Refolding of by rapid dilution

The denatured protein (for both WT and mutant) was centrifuged at 137000 g at 4°C for 20 minutes. The soluble fraction was diluted to a final volume of 200 mL of 8M Urea buffer, with 10 mM DTT, 1 mM GSH, 0.1 mM GSSG and 7 mL/L β-mercaptoethanol. The protein was refolded by rapid dilution (20-fold) into 20 mM Tris and the pH value was adjusted to pH 8.0 with HCl (1^{st} day: pH 9.0; 2^{sd} day pH 8.5; 3^{rd} day: pH 8.0). The refolded protein was then left at 4°C without agitation for one month.

The refolding state of the samples was analyzed by analytical-size exclusion chromatography. The sample was injected in a Superdex 200 HR 10/30 column (GE Healthcare), pre-equilibrated with 50 mM sodium phosphate buffer pH 7.4, 100 mM NaCl pH7.4, using a Prominence Shimadzu HPLC system. The column was calibrated with protein standards in the same buffer: aprotinin (6.5 kDa), ribonuclease (13.7 kDa), carbonic anhydrase (29 kDa), ovalbumin (43 kDa), conalbumin (75 kDa), aldolase (158 kDa) and ferritin (440 kDa).

2.4.2.2 Refolding by dialysis

The denatured protein (for both WT and mutant) was centrifuged at 137000 g at 4°C for 20 minutes. Bearing in mind a final refolding concentration of 0.5 mg/mL, the protein was diluted to a volume of 10 mL of 8 M Urea Buffer with 10 mM DTT, 1 mM GSH, 0.1 mM GSSG and 7 mL/L β-mercaptoethanol. The sample was diluted (10 fold) to a concentration of 0.5 mg/mL in 8 M Urea buffer and then the protein was refolded by dialysis and the pH value was adjusted to pH 8.0 (1^{st} day: 20 mMTris/HCl buffer pH 9.0; 2^{nd} day: 20 mM Tris/HCl buffer pH 8.5, 0.4 M Urea; 3^{rd} day: 20 mM Tris/HCl buffer pH 8.0, 0.4 M Urea). The refolded protein was then left at 4°C without agitation for 72 h or for one month.

In both cases, the refolding state of the samples was analyzed by analytical-size exclusion chromatography, using a Superdex 200 HR 10/30 column (GE Healthcare) as described before.

2.4.3 Purification of recombinant 35toDTG-His and 35toATG after refolding

The refolded protein was concentrated using a Amicon Ultra Centrifugal Filter Unit concentrator (Stirred Cell 8200 –Millipore) and, after ultracentrigation at 137000 g for 20 minutes at 4°C, the sample was applied to a size-exclusion chromatography on a Superdex200 prep grade (GE Healthcare), equilibrated with 20 mM Tris/HCl buffer pH 8.0, 0.4 M Urea. The purification was carried out using BioLogic DuoFlow FPLC system (Bio-Rad) at a constant flow of 1 mL/min. Fractions of 10 mL were collected.

The putative monomeric fractions were then applied to an anion exchange Mono-Q 5/50 GL column (Amersham Biosciences). The column was equilibrated and washed with 20 mM Tris/HCl buffer pH 8.0, 0.4 M Urea and the sample was eluted with a linear gradient of NaCl (0 to 1 M of NaCl) in the same buffer. The purification was carried out using a BioLogic DuoFlow FPLC system (Bio-Rad) at a constant flow of 0.75 mL/min. Fractions of 1 mL were collected through the elution steps. The protein was quantified using a NanoDrop ND 1000 Spectrophotometer.

2.5 Proteolytic Activity evaluation

2.5.1 Effect of pH in enzymatic activity

The effect of pH on 35toDTG-His was tested towards the typical aspartic protease subtract (MCA)-K-K-P-A-E-F-F-A-L-K-(DNP). The assay was carried out with 5 µM of substrate, 50 µL of the protein sample and 50 mM sodium acetate buffer, 100 mM NaCl buffer at pH 3.0, 3.5, 4.0, 4.5, 5.0, and 6.0, or 50 mM sodium citrate buffer, 100 mM NaCl at pH2.5, to a final volume of 200 µL. The reaction was performed in 96-well plates read by a Spectramax Gemini EM fluorimeter (Molecular Devices) set to λ_{ex}=328nm and to λ_{em}=393nm, at 37°C for 3 h.

2.5.2 Digestion of oxidized Insulin β-chain

To evaluate proteolytic activity towards oxidized insulin β-chain, one volume of enzyme was incubated with 3.5 volumes of oxidized insulin β-chain in 100 mM sodium acetate buffer pH 3.5 at 1 mg/mL. Digestion was carried out for 24 h at 37°C, in a water bath. Reaction was stopped by adding TFA (final concentration 0.6%) to the resulting peptides. After centrifugation at 12000 g for 5 minutes, at room temperature, the soluble fragments were separated by a reverse-phase HPLC in a KROMASIL C18 column (Teknokroma) using a Prominence Shimadzu HPLC system, pre-equilibrated with 0.1 % TFA. The elution was carried out by a linear gradient of acetonitrile (0 to 80%) in 0.1 % TFA at a flow rate of 1 mL/min. Elution of peptides was monitored at 220 nm.

2.5.3 Assays in the presence of inhibitors

To test the effect of inhibitors in proteolytic activity, 3.4 µmg of enzyme were incubated with different inhibitors (inhibitors and final concentrations used in the assay are listed in Table 3) in 50 mM sodium acetate buffer pH 3.5, 100 mM NaCl, to a final volume of 200 µL, for 5 minutes at room temperature. After adding substrate to the final concentration of 5 µM, the reaction was carried on 96-well plates read by a

Spectramax Gemini EM fluorimeter (Molecular Devices) set to λ_{ex}=328nm and to λ_{em}=393nm, 37°C for 3 h.

Table3. Inhibitor used in activity assays, final concentration and solvent used for storage.

Inhibitor	Final Concentration
Pepstatin (EtOH)	1µM
Bestatin (EtOH)	10µM
EDTA (H_2O)	5mM
E64 (H_2O)	10µM
Pefabloc (H_2O)	1mM
DTT (H_2O)	1mM
GSH (H_2O)	2mM
GSSG (H_2O)	2mM
NADH (H_2O)	3mM
NADP (H_2O)	2mM
ATP (H_2O)	1mM
ADP (H_2O)	1mM
GTP (H_2O)	1mM
CTP (H_2O)	1mM
datp (H_2O)	1mM
dGTP (H_2O)	1mM
dCTP (H_2O)	1mM
dTTP (H_2O)	1mM

2.6 SDS-PAGE and Western-blot

Samples were denatured by incubation at 95°C for 10 minutes with loading buffer (0.35 M Tris-HCl/0.28% SDS buffer pH 6.8, 30% Glycerol, 10% SDS, 0.6 M DTT and 0.012% Bromophenol Blue). Protein samples were then separated by SDS-PAGE using 12.5% polyacrylamide gels. The gels were run in a MiniProtean 3 system (Bio-Rad) in running buffer (100 mM Tris, 100 mM Bicine, 0.1% SDS) at room temperature at 150V. Gels were stained with Coomassie Brilliant Blue (50% methanol, 10% acetic acid and 0.2 % Brilliant Blue R) and then were incubated with 25% methanol and 5% acetic acid for removal of excess dye.

For Western Blot, the proteins were separated by SDS-PAGE as previously described. Proteins were then transferred to a PVDF membrane, previously activated with methanol. Electrotransference was carried on overnight at 40V at 4°C, in 25 mM Tris, 192 mM Glycine and 20% methanol, using a Trans-BlotR Electrophoretic Transfer cell (Bio-Rad).

PVDF membranes were then blocked for 1 hour with 10 mM Tris pH 8.0, 150 mM NaCl, and 0.1% Tween 20 (TBS-T buffer) containing 5% milk. Membranes were incubated with primary antibody (anti-His (dilution 1:10000) or anti-NP_181826 (dilution 1:20000) for 1 hour or anti-Strep (dilution 1:2000) for 2 hours) (GenScript) in TBS-T buffer containing 05% milk. Membranes were washed at least 7 times for 5 minutes in TBS-T buffer containing 0,5% milk. Membranes incubated with anti-His or anti-Strep were incubated with the secondary antibody Anti-Mouse IgG+IgM alkaline phosphatase linked whole antibody (Amersham Biosciences) with a dilution of 1:10000 in TBS-T with 0.5% milk. On the other hand, membranes incubated previously with anti-NP_181826 were incubated with the secondary antibody Anti-Rabbit IgG+IgM alkaline phosphatase linked whole antibody (Amersham Biosciences) with a dilution of 1:20000 in TBS-T with 0.5% milk. Membranes were revealed with ECF™ substrate (GE Healthcare) in contact with the membrane for 5 minutes and then scanned in a Molecular Imager FX (Bio-Rad).

3. Results and Discussion

3.1 Expression screening of NP_181826 constructs

In order to produce recombinant NP_181826, the coding sequence was chemically synthetized with codon usage optimized for heterologous expression in *E. coli*.

NP_181826 coding sequence was then amplified by PCR, resulting in two amplicons: ΔSPTP and 35toDTG, both with a 5'-end NdeI and a 3'-end XhoI restriction sequences. ΔSPTP sequence does not contain the putative signal peptide neither the putative transit peptide. This sequence could represent a precursor form of NP_181826 that still has the putative DNA-binding domain. This construct has particular interest because it can reveal some features regarding the possible influence of DNA in NP_181826 proteolytic activity/stability, as well as on NP_181826 processing (Fig6).

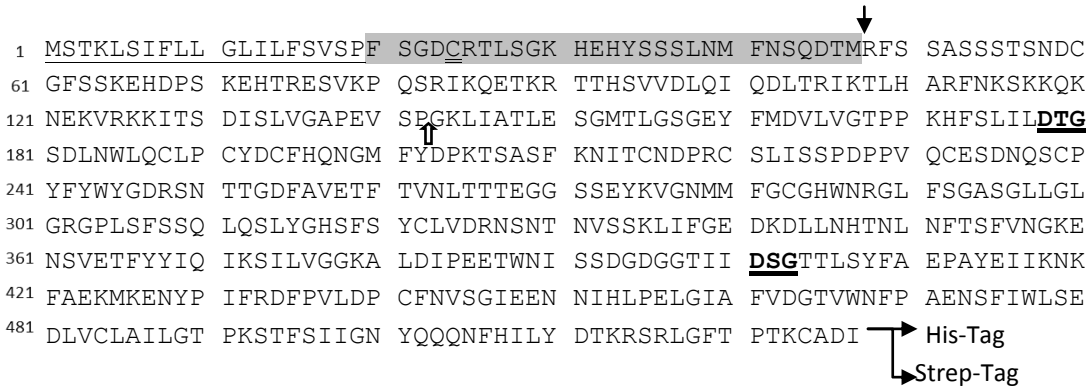

Figure 6. Amplicons amino acid sequence. The signal peptide is underlined; the putative transit peptide is highlighted in gray; catalytic triads are in bold underlined; the beginning of the ΔSPTP amplicon is highlighted by the black arrow and the beginning of 35toDTG is highlighted by the white arrow. ΔSPTP ended with and His-tag and 35toDTG ended with an His-tag or with a Strep-tag.

35toDTG sequence lacks the first 142 amino acids, representing a putative mature and active form of NP_181826 by starting 35 amino acids before the first catalytic triad, DTG.

In general, the N-terminal region of these proteases appears to be important for their maturation processes and regulation of their activity. And so, bearing in mind that the N-terminal domain may be relevant to NP_181826, the chosen restriction sequences allow cloning of these constructs in frame with a His-tag or with a Strep-tag at their C-terminus. The presence of a His-tag or a Strep-tag will help in the purification process

and it will allow its detection through the use of an anti-His antibody or anti-Strep antibody.

The constructs were subcloned in a pET23a vector, with a T7*lac* promoter. In order to express these NP_181826 constructs, several conditions were tried in a small-scale expression system.

3.1.1 Expression screening of recombinant ΔSPTP

The high content of cysteine residues suggests that disulfide bond may have an important role in the structure/stability and proteolytic activity of NP_181826. Several procedures are commonly used to promote the disulfide bond formation (Nozach et al. 2013).

Exporting proteins to the periplasmic space is one example of these procedures. The oxidizing environment and the disulfide bond formation systems can allow the correct folding of the proteins (Nozach et al. 2013). However, secretion of protein to the periplasm can be problematic, mostly because of an insufficient capacity of the translocation machinery; it often leads to low levels of protein, thanks to its limited volume and also to the formation of incorrect disulfide bonds (Berkmen 2012; Nozach et al. 2013). The protein can also be expressed as inclusion bodies, but its refolding is a complex process that can result in a low quantity of protein with the right folding. In alternative, these proteins can be expressed in the cytoplasm, creating conditions that can lead to correct disulfide bonds. The cytoplasm of most prokaryotes has a reducing environment, promoting the reduction of the disulfide bonds (Hatahet et al. 2010). Because of this limitation several *E. coli* strains have been engineered. These strains have a total or partial disruption of one or both pathways that maintains the natural reducing cytoplasmic environment (Hatahet et al. 2010). These commercial strains, such as Origami, also present an additional mutation on *ahpC* gene, in order to increase the growth rate that is reduced because of the disruption of the reducing pathways (Nozach et al. 2013; Hatahet et al. 2010). Although these strains can produce a higher level of disulfide bonded proteins that wild-type strains, the yields of many

proteins obtained are low. Also, the disulfide bond formed may not be the correct disulfide bonds (Nguyen et al. 2011).

Ruddock and colleagues (Hatahet et al. 2010) have shown that co-expression of a sulfhydryl oxidase, such as Erv1p, plus a disulfide isomerase, such as DsbC, can promote an efficient formation of disulfide bonds in the cytoplasm (Hatahet et al. 2010; Nguyen et al. 2011). The co-expression of Erv1p proved to be more efficient that commercial strains, increasing the yields of active protein (Nguyen et al. 2011). Based on these evidences, this was one of the strategies followed in this work.

To obtain higher levels of recombinant ΔSPTP in the soluble fraction, several conditions were tested. Different expression temperatures, *E. coli* strains, with or without Erv1p and DsbC, different media, different IPTG concentrations and different times of expression were used in these small-scale expression screening. The screening was carried out in 15 mL of the selected medium.

The ΔSPTP - pET23a expression construct was transformed in C41, BL21 Star and Origami *E. coli* strains, with and without Erv1p oxidase and DsbC isomerase. The protein expression was tested against two different culture media (LB and TB), two different IPTG concentrations (0.1mM and 0.05mM), two temperatures (37°C and 20°C) and two different expression times (3h and overnight expression).

After expression, samples of each condition were harvested, lysed and soluble and insoluble fractions were analyzed by SDS-PAGE followed by Western blot. Detection was made using an anti-His antibody, which interacts with the C-terminus His-tag (Fig.7). The expected molecular weight of ΔSPTP was 53 kDa.

As it can be seen, in the large majority of conditions tested, the presence of the protein of interest in the soluble fraction is always very low. In fact, the presence of the oxidase seems to slightly increase the yields of soluble protein in only a few conditions.

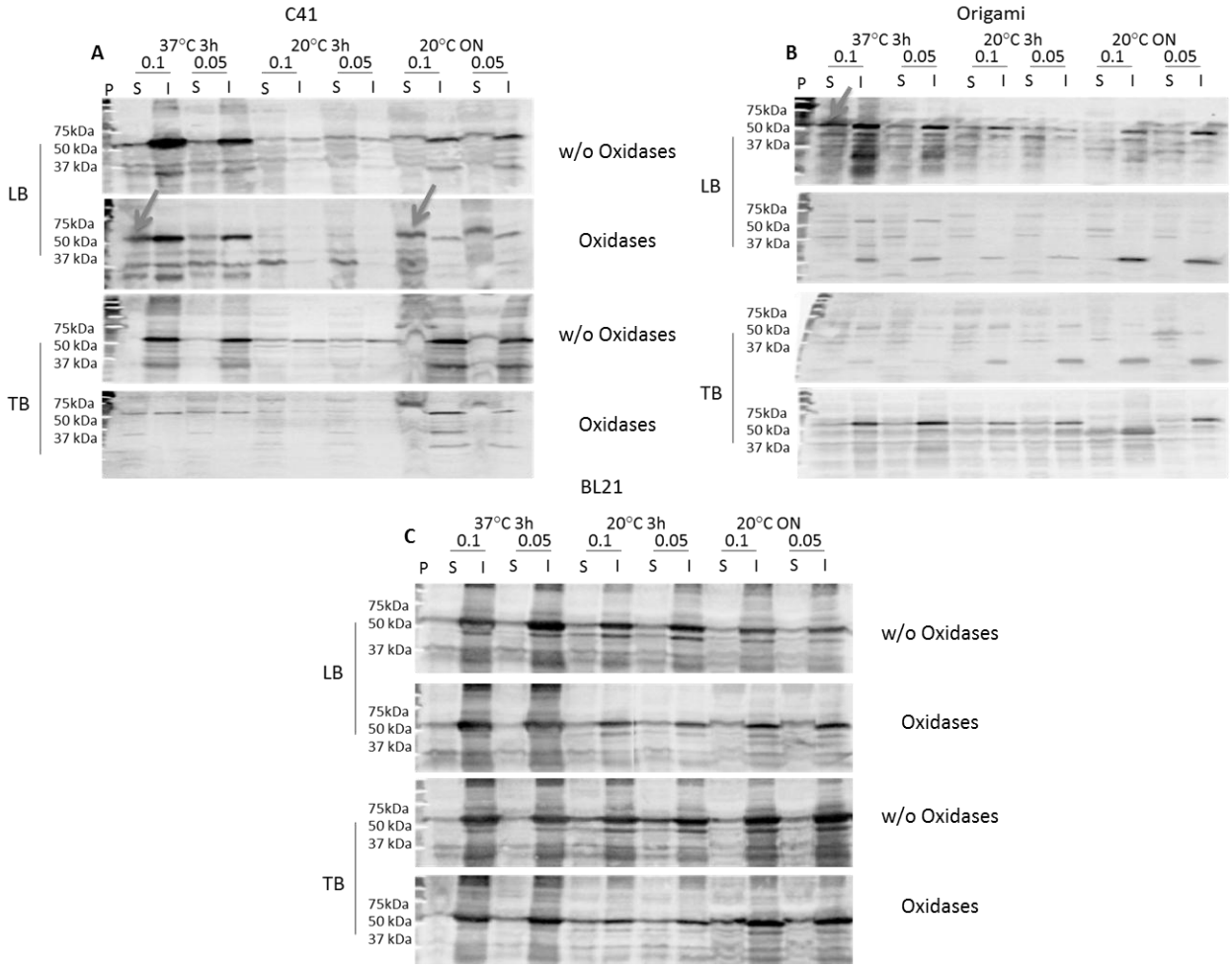

Figure 7. Expression screening of ΔSPTP. Western blot analysis of soluble and insoluble fractions of ΔSPTP expression with anti-His antibody. The protein was expressed in three different strains (C41 [**A**], Origami [**B**] and BL21 star[**C**]) co-transformed with Erv1p and DsbC (designated as Oxidase) or without Erv1p and DsbC (designated as w/o Oxidase), in two different media, two IPTG concentrations, two temperatures and two expression times. **S** stands for soluble fraction and **I** stands for insoluble fraction. The best conditions are highlighted with arrows. P is the molecular weight marker.

Even though the expression levels for soluble protein are still very low, two conditions with C41 cells co-expressing Erv1p may be selected for scale-up expression studies: 1) 37°C for 3h, with 0.1 mM IPTG in LB and 2) 20°C overnight (ON) with 0.1 mM IPTG, in LB. In the first condition, there is a higher accumulation of inclusion bodies. This insoluble fraction may have a negative influence in the soluble protein, promoting its aggregation and, therefore, decreasing the success rate of the purification process.

In the second condition, the expression at lower temperature and for more time resulted in an increase in the soluble fraction and a considerable decrease of the insoluble fraction. Therefore, this should be an interesting condition for expressing this construct. Also, one condition with Origami, (37°C, 3h, with 0.1 mM IPTG, in LB without oxidases) seems to be a good condition for the expression of ΔSPTP. However, this condition also produces a lot of inclusion bodies, which can make the purification process more difficult.

BL21Star seems to preferentially accumulate the protein as inclusion bodies under all tested conditions, with either TB or LB and with or without the oxidase.

Although it was possible to select three conditions for expression of ΔSPTP, due to the higher expression levels observed for the construct 35toDTG-His when compared to these (see next section), the work focused only on the expression of recombinant 35toDTG and no scale-up expressions were further tested for ΔSPTP.

3.1.2 Expression screening of recombinant 35toDTG-His

An expression screening was also made to 35toDTG-His expression construct. As was previously described, different expression temperatures (37°C and 20°C), *E. coli* strains (C41, BL21 Star, Origami), with or without Erv1p and DsbC , different media (LB and TB), different IPTG concentrations (0.1 mM and 0.05 mM) and different times of expression (3 h or ON) were used to do the expression screening. After expression, samples were harvested, lysed and soluble and insoluble fractions were analyzed by SDS-PAGE followed by Western blot. Detection was made using an anti-His antibody (Fig.8). The expected molecular weight of 35toDTG-His was 43 kDa

For this construct it was possible to identify a higher number of conditions with increased yields of soluble protein. The presence of Erv1p and DsbC appear to promote protein solubility under selected conditions. This has been described by Nguyen and co-workers (2010) for Origami. C41 cells co-expressing the oxidases at 37°C, in TB or LB, with both 0.1 mM and 0.05 mM IPTG, display increased levels of soluble protein.

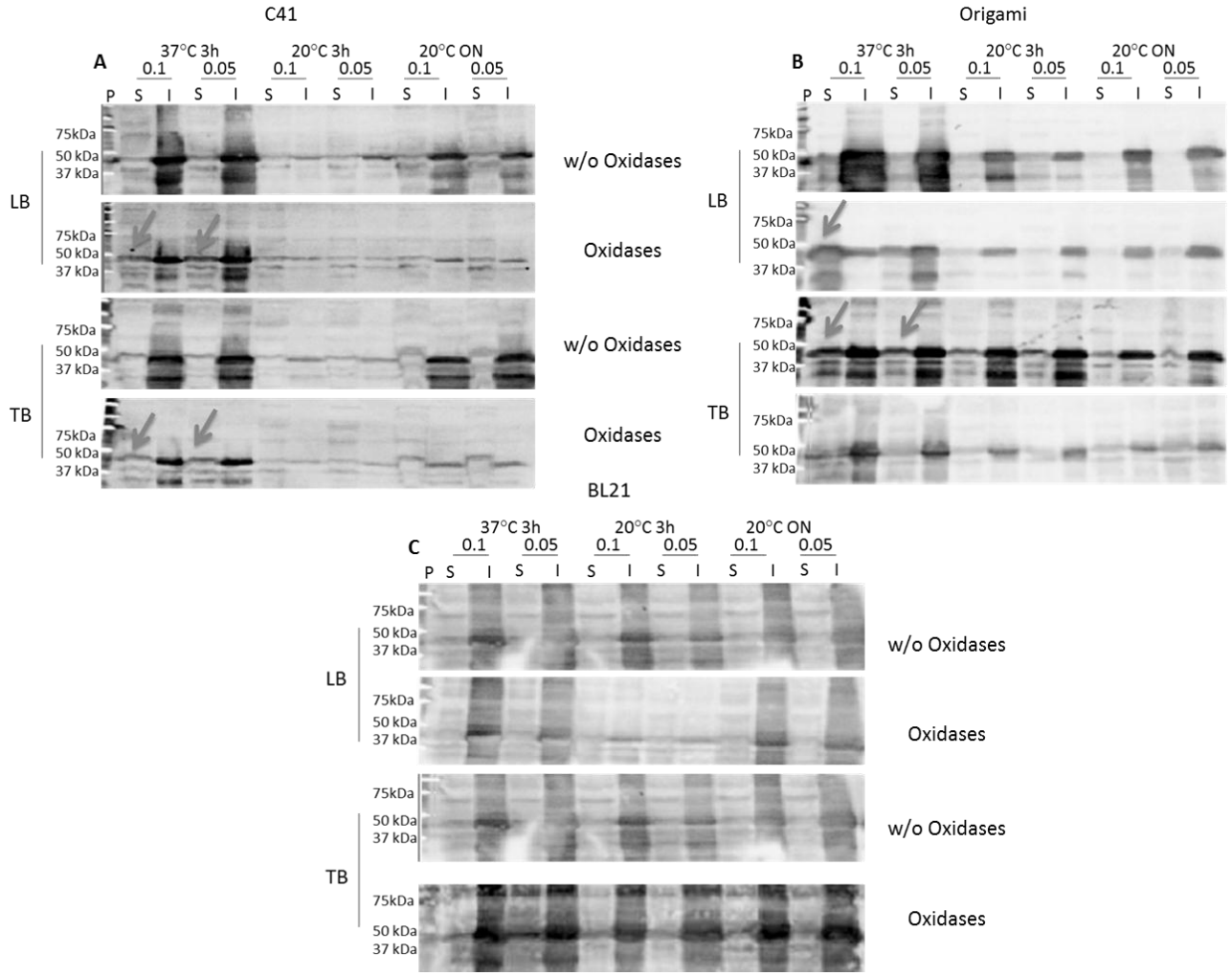

Figure 8. Expression screening of 35toDTG-His. Western blot analysis of soluble and insoluble fractions of 35toDTG-His expression with anti-His antibody. The protein was expressed in three different strains (C41 [**A**], Origami [**B**] and BL21 star[**C**]) co-transformed with Erv1p and DsbC (designated as Oxidase) or without Erv1p and DsbC (designated as w/o Oxidase), in two different mediums, two IPTG concentrations, two temperatures and two expression times. **S** stands for soluble fraction and **I** stands for insoluble fraction. The best conditions are highlighted with arrows. P is the molecular weight marker.

Also, with Origami cells at 37°C with 0.1 mM IPTG in LB medium, co-expressing the oxidases, and with 0.1 mM and 0.05 mM IPTG in TB medium, without the oxidases, there is an increased accumulation of soluble protein. Similar to ΔSPTP, BL21Star strain seems to promote 35toDTG-His expression as inclusion bodies in all tested conditions.

Comparing the expression screening for both constructs it is easy to conclude that each protein is expressed in a unique way, even under the same conditions. Also, the

presence of the oxidase has a different influence in the expression of each construct. The presence or absence of Erv1p and DsbC is more significant when expressing 35toDTG-His, being the increase of soluble protein, compared to the insoluble fraction for each condition, more clear.

Because the results suggest that 35toDTG-His is more expressed as soluble protein than ΔSPTP, it was decided to continue this work by expressing 35toDTG-His_NP under the following conditions: Origami cells co-expressing oxidases, 37°C (3h) with 0.1 mM IPTG, in LB. This condition shows a higher expression of soluble protein in comparison to accumulation in inclusion bodies.

3.2 Expression and Purification of 35toDTG-His as soluble protein

After choosing the best condition to express 35toDTG-His as a soluble protein, the expression process was scaled-up (2 L) to obtain more protein for optimization of purification process and biochemical characterization. Co-transformed Origami cells with Erv1p, DsbC and the 35toDTG-His pET23a construct were cultivated in 2 L LBAmp/Cam in Fernbach flasks at 37°C with constant rotation to an OD600nm of 0.4 and then pre-induced with 0.5 %w/v Arabinose. After 30 minutes, protein expression was induced with 0.1 mM IPTG. The culture was incubated at 37°C with constant rotation for 3 h.

After expression, the cells were harvested and the culture pellet was resuspended in 10 mL per 1 L of culture with 20 mM sodium phosphate buffer pH7.5, 10 mM Imidazole, 500 mM NaCl and then frozen at -20°C.

3.2.1 Recombinant 35toDTG-His purification with IMAC-Ni^{2+} and IMAC-Co^{2+}

Harvested cells expressing 35toDTG-His were thawed at room temperature, lysed and the soluble protein fraction obtained was divided and applied to two HisTrap HP 5 mL IMAC columns charged with Ni^{2+}. The two affinity columns were independently washed with 20 mM phosphate buffer pH 7.5, 10 mM Imidazole, 500 mM NaCl, for avoiding unspecific interactions. Protein was eluted with 20 mM phosphate buffer pH

7.5, 500 mM NaCl in 50 mM, 100 mM and 500 mM Imidazole elution steps. Imidazole competes with the His-tag for interaction with immobilized nickel ions, promoting the elution of the protein. The purification step was carried out using a BioLogic DuoFlow FPLC system (Bio-Rad) at a constant flow of 5 mL/min. Protein elution was monitored by detecting A280nm. The chromatogram obtained is shown in Fig. 9.

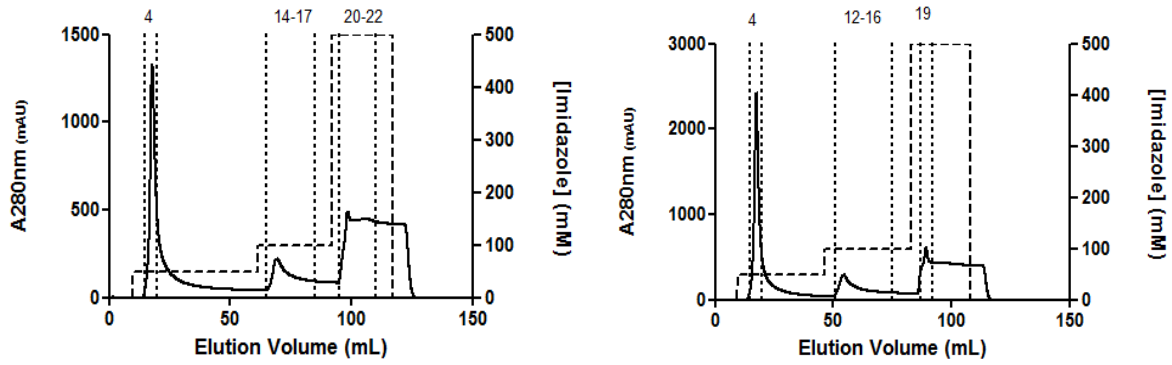

Figure 9. Purification of recombinant 35toDTG-His by IMAC-Ni^{2+} chromatography. The column was previously equilibrated with 20 mM phosphate buffer pH7.5, 10 mM Imidazole, 500 mM NaCl and washed with the same buffer. The protein was eluted with 20 mM phosphate buffer pH7.5, 500 mM NaCl in three elution steps with increasing imidazole concentration, as it is shown (50 mM, 100 mM and 500 mM). The purification was carried out at a constant flow of 5 mL/min. The collected fractions for each IMAC-Ni^{2} highlighted above (14-17 and 12-16).

As it can be seen, a large amount of protein was eluted in the first step of Imidazole (50 mM). Although 35toDTG-His was also eluted in this step (see Fig.10), most of the proteins present in this fraction are proteins that had an unspecific interaction with the column. The protein of interest was eluted with 100 mM and 500 mM Imidazole. Given the low protein levels, collected fractions from the 100 mM imidazole step of the two IMAC columns were polled together so that a higher amount of protein could be obtained.

Samples from the three elution steps were analyzed by SDS-PAGE and stained with Coomassie Blue (Fig.10).

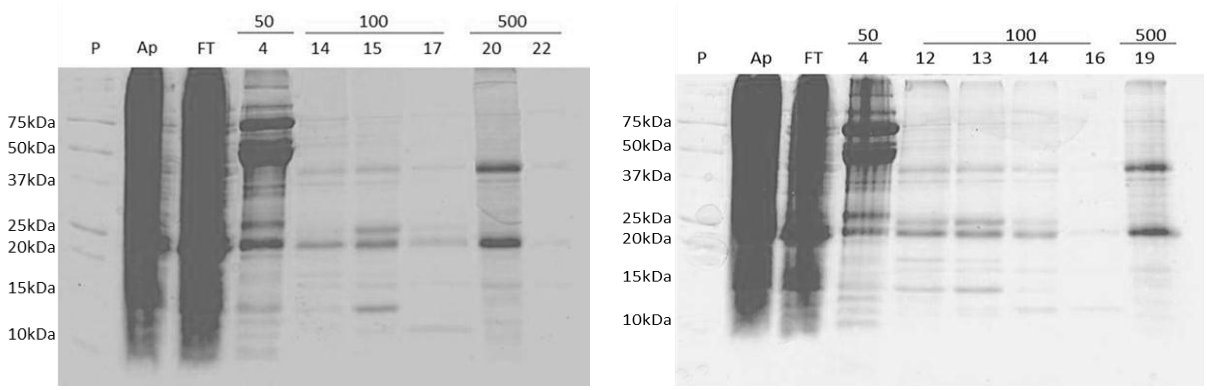

Figure 10. SDS-PAGE analysis of the factions collected in the purification of recombinant 35toDTG-His by IMAC-Ni^{2+} chromatography. Fraction numbering corresponds to the numbers displayed in the chromatograms from Fig.9; 50, 100 and 500, correspond to imidazole elution steps; AP stands for the total soluble extract applied to the column and FT to the protein not bound to the column (flow through). P is the molecular weight marker.

The theoretical molecular weight of 35toDTG-His is 43kDa. 35toDTG-His was eluted in all the Imidazole steps and the protein was co-purified with other proteins. The elution in all steps and the presence of these contaminants may be a signal that the protein is not correctly folded and these co-purified proteins probably correspond to chaperones present in *E. coli*.

Although the fractions from the 500 mM Imidazole step contained a larger amount of protein, this protein appeared to be aggregated as no protein was recovered in subsequent purification steps (data not shown).

The factions collected in the second elution step were subjected to further purification. To concentrate the protein and obtain more pure samples, the 100 mM fractions were pooled and dialyzed against 20 mM phosphate buffer pH 7.5, 100mM NaCl at 4°C overnight. The dialyzed sample was applied to HisTrap HP 1 mL columns (Amersham Biosciences) charged with Co^{2+}. The column was also washed with 20 mM phosphate buffer pH 7.5, 10 mM Imidazole, 500 mM NaCl and the sample was eluted with 20 mM phosphate buffer pH 7.5, 500 mM NaCl in 50 mM, 100 mM 500 mM Imidazole elution steps. Like before, the Imidazole will compete with the His-tag to interact with the Cobalt ions. The purification step was carried out at a constant flow of 1 mL/min. Fractions of 1 mL were collected through the elution steps (Fig.11).

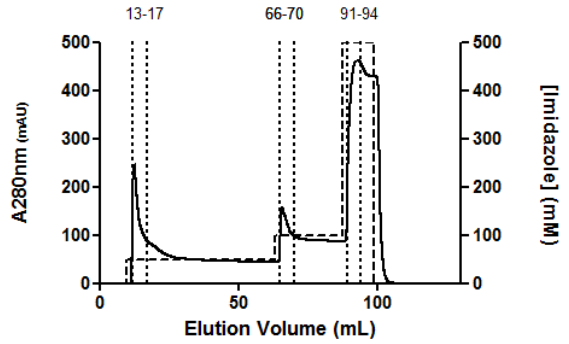

Figure 11. Purification of polled sample from IMAC-Ni^{2+} chromatography containing 35toDTG-His by IMAC-Co^{2+} chromatography. The column was previously equilibrated with 20 mM phosphate buffer pH 7.5, 10 mM Imidazole, 500 mM NaCl and washed with the same buffer. The protein was eluted with 20 mM phosphate buffer pH7.5, 500 mM NaCl in three elution steps with increasing imidazole concentration, as it is shown (50 mM, 100 mM and 500 mM). The purification was carried out at a constant flow of 1 mL/min. The collected fractions are highlighted above (13-17, 66-70, 91-94).

Proteins that interact nonspecifically with the column were mainly eluted in the first elution step. 35toDTG-His has the similar behavior previously seen in IMAC-Ni^{2+} columns. The protein was eluted in all the Imidazole steps as well. The samples were analyzed by SDS-PAGE, stained with Coomassie Blue, and Western Blot with an anti-His antibody (Fig.12).

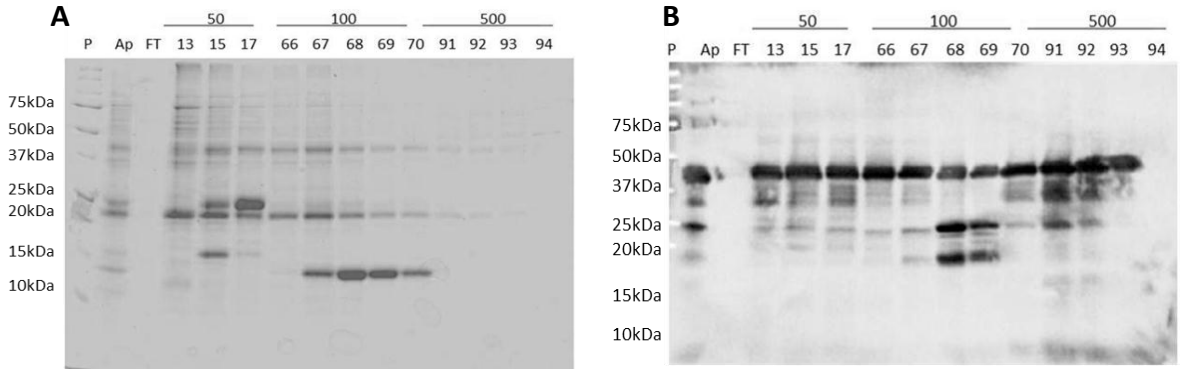

Figure 12. SDS-PAGE [A] and Western Blot [B] analysis of the factions collected from IMAC-Co^{2+}. The numbers correspond to the fraction numbers from the chromatogram in Fig. 11; 50, 100 and 500, correspond to imidazole elution steps; AP stands for the total soluble extract applied and FT to the protein not bound to the column. SDS-PAGE was stained with Coomassie Blue. Antibody anti-His-tag (1:10000) was used to identify 35toDTG-His in the Western blot. P corresponds to molecular weight markers.

Most of the protein of interest is eluted in second elution step, though the protein levels are still low. Also, this purification step wasn't able to clean the fraction from contaminants, suggesting again that these proteins present in the samples must interact with 35toDTG-His and be co-purifying with it.

Even so, the more concentrated fractions 66 and 67 were pooled and dialyzed against 20 mM Tris/HCl buffer pH 7.5 with 50 mM NaCl, at 4°C overnight, and then used to screen for enzymatic activity. Although several substrates were tested, the enzyme only showed activity with the typical AP substrate (data not shown). So the assays were performed at different pH values, using the typical AP substrate (MCA)-K-K-P-A-E-F-F-A-L-K-(DNP). The pH profile of obtained is shown in Fig. 13.

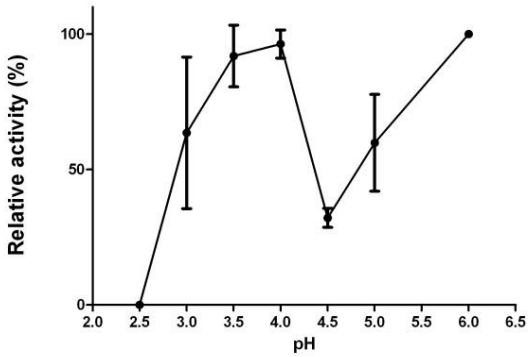

Fig13. pH profile of recombinant 35toDTG-His enriched fractions purified with IMAC-Ni^{2+} and IMAC-Co^{2+}. The effect of pH in 35toDTG-His activity was assayed using the typical AP substrate [(MCA)-K-K-P-A-E-F-F-A-L-K-(DNP)] and several buffers with different pH values (50 mM sodium acetate buffer, 100 mM NaCl at pH 3.0, 3.5, 4.0, 4.5, 5.0, 6.0, or 50 mM sodium citrate buffer, 100 mM NaCl buffer at pH 2.5). The reaction was performed at 37°C for 3h and negative controls were made for each pH value using only sample buffer, substrate and the assay buffers.

The assays show a tendency for an optimum pH value at pH 4.0 towards this typical aspartic protease substrate. An acidic optimum pH value is usually seen in aspartic proteases. The results also showed the presence of proteolytic activity at pH 6.0, most likely because of the presence of contaminants in the sample.

In order to have cleaner samples and to obtain more protein that allows a better biochemical characterization of the enzyme a different purification strategy was used.

3.2.2 Recombinant 35toDTG-His purification with IMAC-Ni^{2+} and Anionic Exchange Chromatography

Harvested cells expressing 35toDTG-His were lysed and the soluble protein fraction was applied to a Histrap HP 5 mL IMAC column charged with Ni^{2+}.

The column was again washed with 20 mM phosphate buffer pH 7.5, 10 mM Imidazole, 500 mM NaCl and protein was eluted with 20 mM phosphate buffer pH 7.5, 500 mM NaCl in 50 mM, 100 mM and 500 mM Imidazole elution steps. The purification step was carried out as described before at a constant flow of 5 mL/min. Fractions of 5mL were collected (Fig.14).

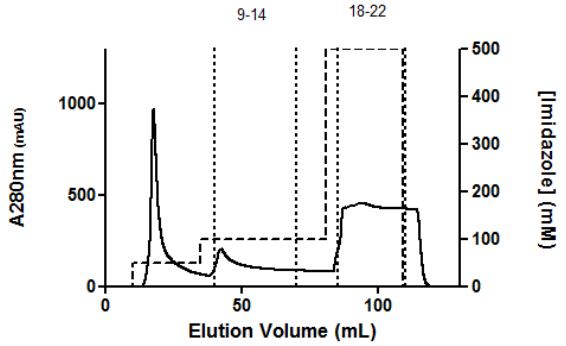

Figure 14. Purification of recombinant 35toDTG-His by IMAC-Ni^{2+} chromatography. The column was previously equilibrated with 20mM phosphate buffer pH 7.5, 10 mM Imidazole, 500 mM NaCl and washed with the same buffer. The protein was eluted with 20 mM phosphate buffer pH 7.5, 500 mM NaCl in three elution steps with increasing imidazole concentration (50 mM, 100 mM and 500 mM). The purification was carried out at a constant flow of 5 mL/min. The collected fractions are highlighted above (9-14, 18-22).

As it was seen before, though 35toDTG-His is eluted in all three elution steps (see Fig.15). Samples from the three elution steps were analyzed by SDS-PAGE stained with Coomassie Blue. These samples were also analyzed by SDS-PAGE followed by Western blot. Anti-His-tag (1:10000) was used to identify 35toDTG-His in the blot (Fig.15).

In a similar process to the one described before, fraction from 100 mM Imidazole were pooled and dialyzed against 20mM Tris/HCl buffer pH 7.5 with 100mM NaCl at 4°C overnight. The dialyzed fraction was applied to an anion exchange Mono-Q 5/50 GL column (Amersham Biosciences).

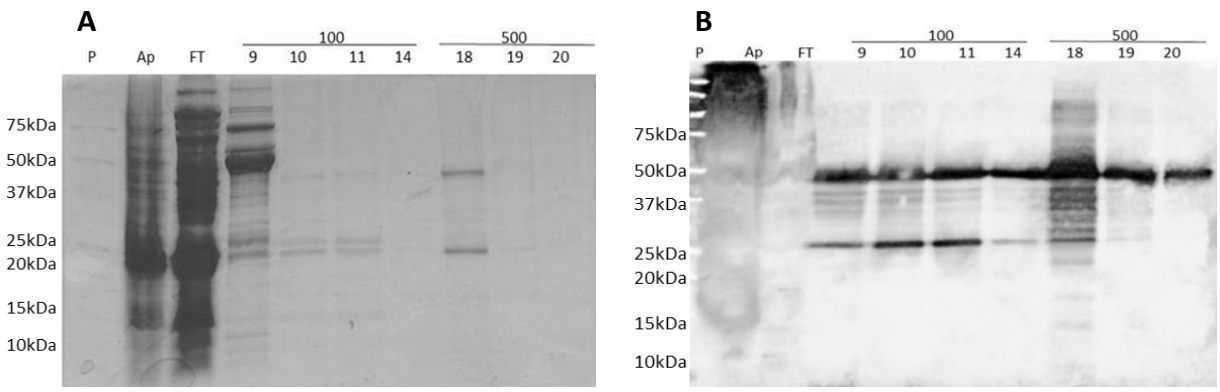

Figure 15. SDS-PAGE [A] and Western Blot [B] analysis of the fractions collected from IMAC-Ni^{2+}. The numbers correspond to the fraction number from the chromatogram shown in Fig.14; AP stands for the total soluble extract applied and FT to the protein not bounded to the column; 50, 100 and 500, correspond to imidazole elution steps. SDS-PAGE was stained with Coomassie Blue. Antibody anti-His-Tag (1:10000) was used to identify 35toDTG-His in the Western blot. P corresponds to molecular weight markers.

35toDTG-His has a theoretical pI of 4.98. So the column was equilibrated and washed with 20 mM Tris/HCl buffer pH 7.5. 35toDTG-His have a negative net charge at pH 7.5, interacting with the cationic matrix of the column. The sample was eluted with a linear gradient of NaCl (0 to 1 M of NaCl) in the same buffer. Salt ions compete with the protein to interact with the column's positive matrix, promoting the elution of the proteins. Proteins with stronger interaction with the matrix need a higher concentration of NaCl to be eluted. The purification step was carried out using a BioLogic DuoFlow FPLC system (Bio-Rad) at a constant flow of 0.75 mL/min. Protein elution was monitored by detecting A280nm. Conductivity increases as NaCl concentration increases as well. So conductivity was used to monitorize the increase of salt concentration. Fractions of 1 mL were collected through the elution steps (Fig. 16). As observed with the IMAC purification, 35toDTG-His was eluted through the entire linear gradient (see Fig.17).

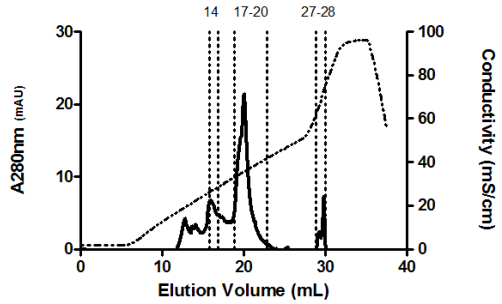

Figure 16. Purification of polled sample from IMAC-Ni^{2+} chromatography containing 35toDTG-His by anionic exchange chromatography. The MonoQ column was previously equilibrated with 20 mM Tris/HCl buffer pH7.5 and washed with the same buffer. The protein was eluted with a linear gradient of NaCl (0 to 1 M of NaCl). The purification was carried out at a constant flow of 0.75mL/min.

Also 35toDTG-His was again co-purified with other proteins, approximately with the same molecular weight that the contaminations present in IMCA-Co^{2+}. This suggests that 35toDTG-His may be interacting directly with these proteins.

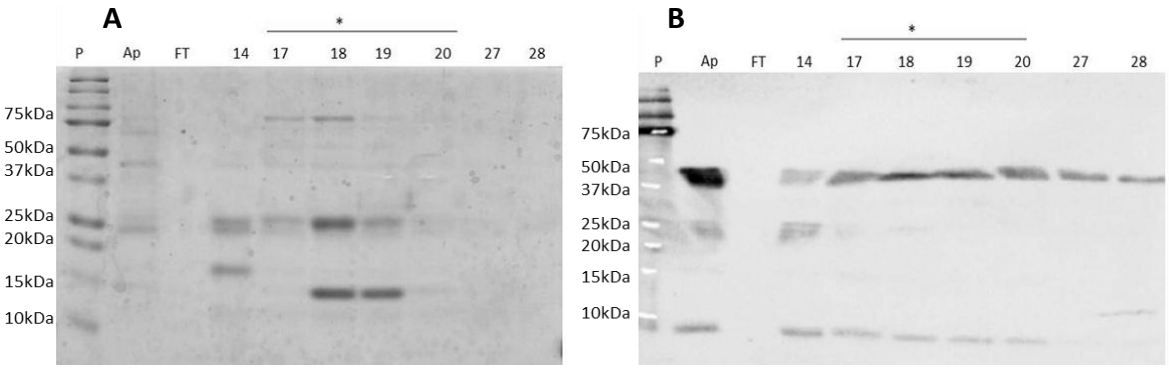

Figure 17. SDS-PAGE [A] and Western Blot [B] analysis of the factions collected from anionic exchange chromatography. The numbers correspond to the fraction number from the chromatogram shown in Fig.16; AP stands for the total soluble extract applied and FT to the protein not bounded to the column. SDS-PAGE was stained with Coomassie Blue. Antibody ant-His-tag (1:10000) was used to identify 35toDTG-His in the Western blot. Fractions 17-20 were used to screen proteolytic activity; fraction with more activity is highlighted (*).P corresponds to molecular weight markers.

Co-purification of 35toDTG-His_NP with other proteins as well as the consistent very low yields of recovered protein suggests that 35toDTG-His may not have the correct folding.

With this purification strategy it was again not possible to improve 35toDTG-His purification.

Even so, fractions from the anionic exchange chromatography were used to screen for proteolytic activity against the typical aspartic protease substrate (MCA)-K-K-P-A-E-F-F-A-L-K-(DNP) in 50 mM sodium acetate buffer pH 4.0, 100mM NaCl buffer (Fig18).

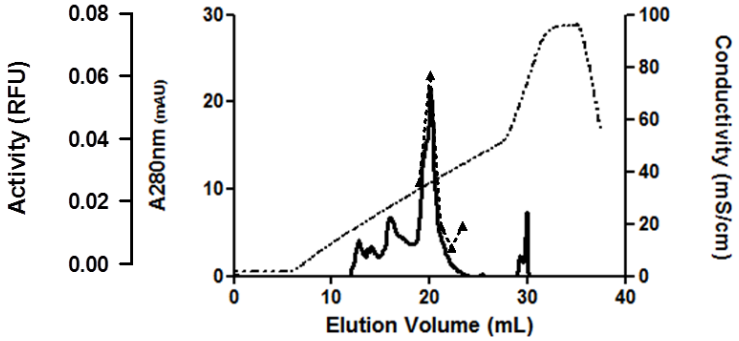

Figure 18. Analysis of the enzymatic activity of fractions eluted by anionic exchange chromatography. Proteolytic activity of fractions eluted from Mono Q was tested towards the typical AP substrate [(MCA)-K-K-P-A-E-F-F-A-L-K-(DNP)] in 50 mM sodium acetate buffer pH 4, 100 mM NaCl. Activity is represented by a discontinuous line in the chromatogram. The reaction was performed at 37°C for 3h and negative controls were made for using only sample buffer, substrate and the assay buffer.

As it can be seen, it was possible to obtain fractions with activity from the central peak (elution volume 20 mL) of the chromatogram, were a higher amount of protein was eluted. Fraction 18, which contain a higher amount of protein and more proteolytic activity, was used obtain a pH profile of 35toDTG-His (Fig.19).

The pH curve showed again an optimum pH value at pH4.0 with the typical AP substrate, corroborating the previous results obtained with the other purification strategy. With this purification method there was no activity at pH6.0. This suggests again that the activity at this higher pH value was probably due the presence of other proteins that we not co-purified by this anion exchange chromatography step.

Although this strategy allowed the elaboration of a pH activity profile, the amount of purified protein and the level of purification were very low and not enough to do a complete biochemical characterization. Therefore, a different strategy was implemented to optimize the purification process.

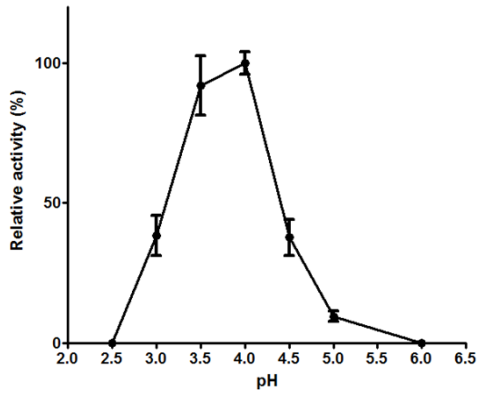

Figure 19. pH profile of recombinant 35toDTG-His enriched fractions purified by IMAC-Ni^{2+} followed by anionic exchange chromatography. The effect of pH in 35toDTG-His activity was assayed using the typical AP substrate [(MCA)-K-K-P-A-E-F-F-A-L-K-(DNP)] and several buffers with different pH values (50 mM sodium acetate buffer with 100 mM NaCl buffer at pH 3.0, 3.5, 4.0, 4.5, 5.0, 6.0, or 50 mM sodium citrate buffer, 100 mM NaCl buffer at pH 2.5. The reaction was performed at 37°C for 3h and negative controls were made for each pH value using only sample buffer, substrate and the assay buffers.

3.2.3 Expression of recombinant 35toDTG-His in Ultra Yield flasks and purification by IMAC-Ni^{2+} and anionic exchange chromatography

Since the amount of purified 35toDTG-His was very low and not enough to perform the biochemical characterization of the enzyme, we decided to use Ultra Yield flasks (BioSilta) to increase the yields of the process. The company BioSilta (http://biosilta.com/) claims that these Ultra Yield flasks have a better oxygenation, due to their structure, allowing an increase in cell densities and allowing cells to grow for longer time, as the cultures are not limited by oxygen rates.

As described before, a co-transformed Origami pre-inoculum was diluted in Ultra Yield flasks (BioSilta), 1:50 (v/v) in LBAmp/Cam medium (1 L per flask), and the cells were incubated at 37°C with constant rotation to an OD_{600nm} of 0.4 and then pre-induced with 0.5% w/v Arabinose; 30 minutes after induction with arabinose, IPTG was added (0.1 mM final concentration) and the cells were incubated at 30°C with constant rotation for 3h.

Then cells from 2 L of expression were harvested, subjected to a frozen/thaw cycle and to mechanical lysis and the soluble protein fraction was applied to a HisTrap HP 5mL

IMAC column charged with Ni^{2+}. The column was washed and protein was eluted as described before. The purification step was carried out in an Akta FPLC system (GE Healthcare) at a constant flow of 5 mL/min. Fractions of 5 mL were collected (Fig.20).

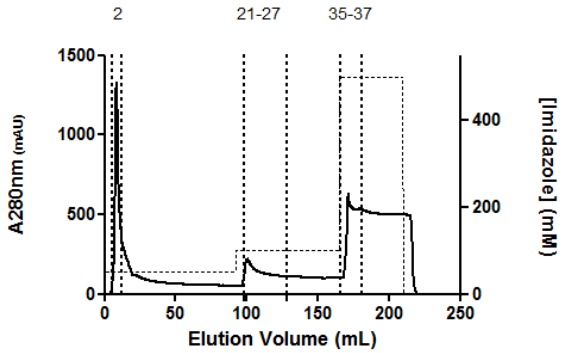

Figure 20. Purification of recombinant 35toDTG-His after expression in Ultra Yield flasks by IMAC-Ni^{2+} chromatography. The column was previously equilibrated with 20 mM phosphate buffer pH 7.5, 10 mM Imidazole, 500 mM NaCl and washed with the same buffer. The protein was eluted with 20 mM phosphate buffer pH 7.5, 500 mM NaCl in three elution steps with increasing imidazole concentration (50 mM, 100 mM and 500 mM). The purification was carried out at a constant flow of 5 mL/min. The collected fractions are highlighted above (2, 21-27, 35-37).

As it can be seen by comparing the chromatogram (Fig. 20) and the analysis by SDS-PAGE and Western blot (Fig.21), a very low yield of purified protein was obtained.

Fractions from 100 mM Imidazole step of elution were pooled and dialyzed against 20mM Tris/HCl buffer pH 7.5, 100 mM NaCl at 4°C overnight.

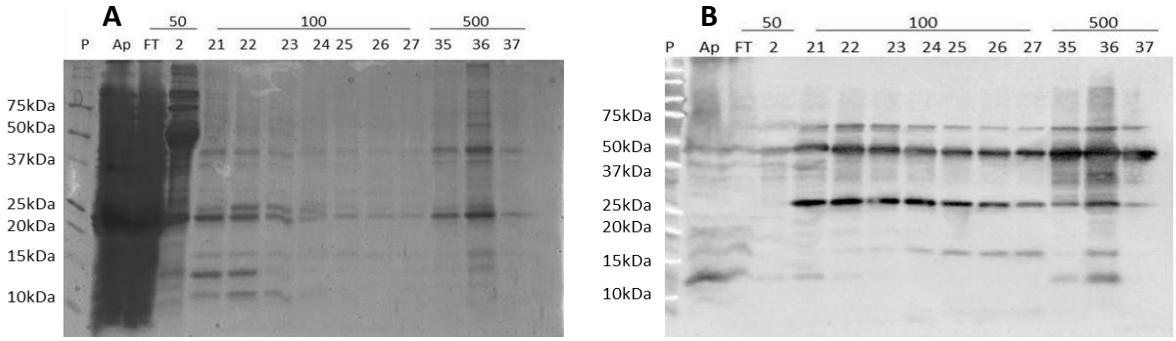

Figure 21. SDS-PAGE [A] and Western Blot [B] analysis of the factions collected from IMAC-Ni^{2}. The numbers correspond to the fraction number from the chromatogram shown in Fig.20; AP stands for the total soluble extract applied and FT to the protein not bounded to the column; 50, 100 and 500, correspond to imidazole elution steps. SDS-PAGE was stained with Coomassie Blue. Antibody ant-His-tag (1:10000) was used to identify 35toDTG-His in the Western blot. P corresponds to molecular weight markers.

The dialyzed pool was applied to an anion exchange Mono-Q 5/50 GL column (Amersham Biosciences) that was equilibrated and washed with 20 mM Tris/HCl buffer pH7.5. The sample was eluted with a linear gradient of NaCl (0 to 1 M of NaCl) in the same buffer. The purification step was carried out using an Akta FPLC system (GE Healthcare) at a constant flow of 0.75 mL/min. Protein elution was monitored by detecting A280nm. Conductivity was used to monitorize the increase of salt concentration. Fractions of 1 mL were collected through the elution gradient (Fig.22).

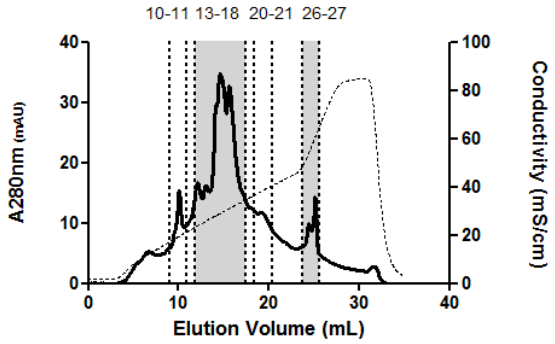

Figure 22. Purification of polled sample from IMAC-Ni^{2+} chromatography containing 35toDTG-His by anionic exchange chromatography. The column was previously equilibrated with 20 mM Tris/HCl buffer pH 7.5 and washed with the same buffer. The protein was eluted with a linear gradient of NaCl (0 to 1 M of NaCl). The purification was carried out at a constant flow of 0.75mL/min. The collected fractions are highlighted above (10-11 and 20-21; 13-18 and 26-27 in grey).

SDS-PAGE and Western blot results show once again that, even with 2 L of expression of 35toDTG-His in Ultra Yields flasks, the purification yield did not increase (Fig.23).

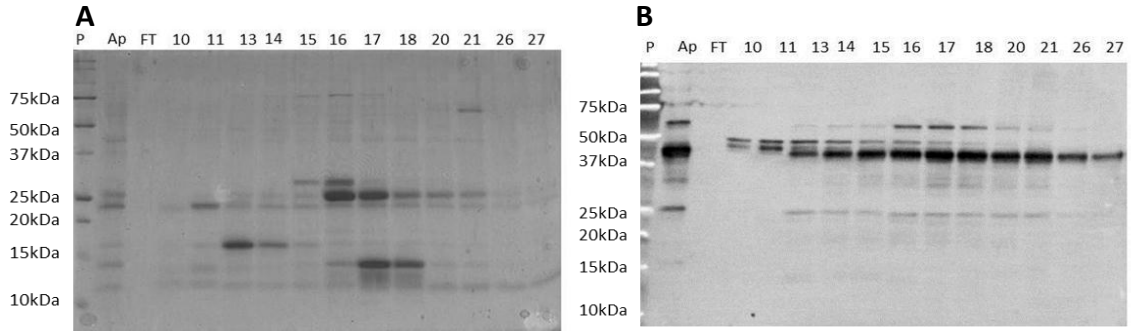

Figure 23. SDS-PAGE [A] and Western Blot [B] analysis of the factions collected from anionic exchange chromatography. The numbers correspond to the fraction number from the chromatogram shown in Fig.22; AP stands for the total soluble extract applied and FT to the protein not bounded to the column. SDS-PAGE was stained with Coomassie Blue. Antibody ant-His-tag (1:10000) was used to identify 35toDTG-His in the Western blot. P corresponds to molecular weight marker.

In fact, the final amount of purified protein is very similar to the amount of protein obtained from an expression in Fernbach flasks and purification with the same process (compare with Fig.17). Based on these results, this strategy for expression of 35toDTG was abandoned.

3.2.4 Expression and Purification of recombinant 35toDTG-Strep

In order to overcome the limitations observed with expression of 35toDTG in the soluble form when fused to a His-Tag, a Strep Tag was used to purify the protein of interest. A Strep tag is an eight-amino-acid peptide (WSHPQFEK) that specifically binds to a modified binding pocket of an engineered streptavidin with improved binding capacity (Schmidt & Skerra 2007; Yan et al. 2006). When comparing His Tag and Strep Tag effectiveness, Strep tag was shown to be more effective, being able to purify highly pure protein in good yields (Yan et al. 2006). Moreover, Strep tag doesn´t appear to interfere with the protein folding or protein function (Schmidt & Skerra 2007).
Based on this, 35toDTG-Strep was amplified using specific primers and it was subcloned into a pET23a expression vector in frame with a C-terminal Strep Tag.

3.2.4.1 Expression screening of recombinant 35toDTG-Strep

To find the best conditions to express recombinant 35toDTG-Strep, different expression temperatures, different IPTG concentrations and different times of expression were used to do small-scale expression tests. The screening was carried out in 15 mL of medium with a dilution of 1:1000 of the correct antibiotic, as described previously.
The pET23a 35toDTG-Strep expression construct was transformed in Origami *E. coli* strains, with Erv1p oxidase and DsbC isomerase. The protein expression was tested against one different culture medium (LB), two different IPTG concentrations (0.1 mM and 0.05 mM), two temperatures (37°C and 20°C) and two different expression times (3h and overnight expression).

After expression, samples of each condition were harvested, lysed and soluble and insoluble fractions were analyzed by SDS-PAGE followed by Western blot. Detection was made using an anti-Strep antibody, which interacts with the C-terminus Strep-tag (Fig.24).

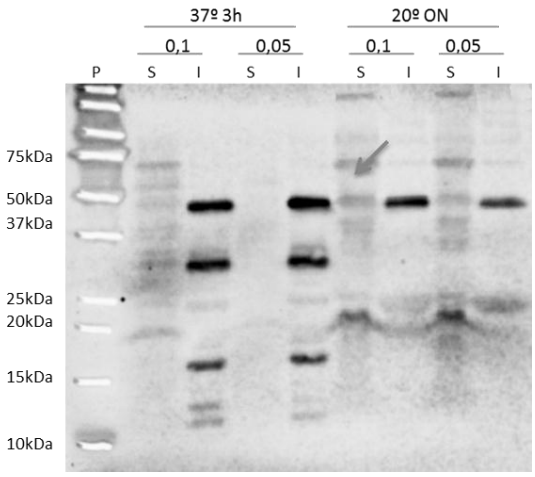

Figure 24. Expression screening of recombinant 35toDTG-Strep. The protein was expressed in Origami co-transformed with Erv1p and DsbC, in LB, with different IPTG concentrations (0.1 mM; 0.05 mM), temperatures (20°C and 37°C) and expression times (3 hours; overnight). **S** stands for soluble fraction and **I** stands for insoluble fraction. The best condition is highlighted with an arrow. P corresponds to the molecular weight marker

At 37°C, 35toDTG-Strep is mostly expressed as inclusion bodies with both IPTG concentrations. To improve the amount of soluble protein, an overnight expression at a lower temperature was performed. The expression under these conditions increased the amount of protein in the soluble fraction. From the conditions tested, expressing 35toDTG-Strep at 20°C overnight with 0.1 mM IPTG showed to be the best condition to express 35toDTG-Strep as soluble protein.

3.2.4.2 Expression in Ultra Yield flasks and Purification of Recombinant 35toDTG-Strep with a Strep Tag affinity chromatography

Co-transformed Origami with the modified pLysS vector and the 35toDTG-Strep pET23a construct were pre-inoculated in an Erlenmeyer flask with LBAmp/Cam medium at 37°C with constant rotation overnight. To increase the amount of protein, expression was carried out using ultra yield flasks, where 1L LBAmp/Cam was

inoculated with 20 mL of the pre-inoculum. The cells were incubated at 37°C with constant rotation to an OD_{600nm} of 0.4. The cultures were then pre-induced with 0.5%w/v Arabinose. After 30 minutes, protein expression was induced with 0.1mM IPTG. The culture was incubated at 20°C with constant rotation overnight.

After expression, 1L cultures were centrifuged at 6000 g for 20 min at 4°C. 35toDTG-Strep's culture pellet was ressuspended in 10 mL per 1 L of culture with 100 mMTris/HCl buffer pH 8.0, 150 mM NaCl. The ressuspension was also frozen at-20ºC.

Harvested cells were thawed at room temperature, lysed and the soluble protein fraction obtained was filtered through 0.2 µm filters. The protein was then applied to one StrepTrap HP 1 mL column (Amersham Biosciences).

The column was equilibrated and washed with 100 mMTris/HCl buffer pH 8.0, 150 mM NaCl and the sample was eluted with 100 mMTris/HCl buffer pH 8.0, 150mM NaCl with 2.5 mM desthiobiotin in one elution step. Desthiobiotin is a biotin analog that is able to bind to the modified biotin-binding pocket of the engineered streptavidin. So, it will compete with the Strep tags for binding with the matrix of the columns, promoting the elution of the protein (Yan et al. 2006).

The purification step was carried out using an AKTA FPLC system (Ge Healthcare) at a constant flow of 0.75 mL/min. The column was inverted in the elution step. Fractions of 1 mL were collected (Fig. 25).

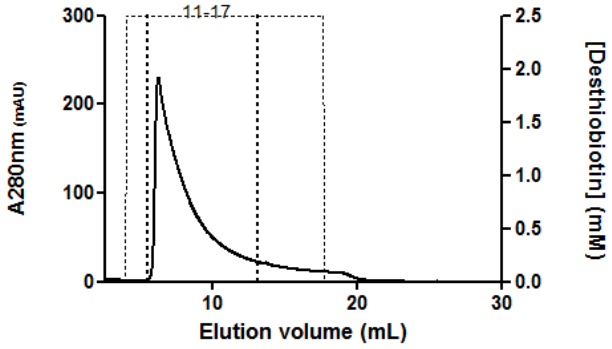

Figure 25. Purification of recombinant 35toDTG-Strep by StrepTrap affinity chromatography. The column was previously equilibrated with 100 mM Tris/HCl buffer pH 8.0, 150mM NaCl and washed with the same buffer. The protein was eluted with 100mM Tris/HCl buffer pH 8.0, 150 mM NaCl, 2.5 mM desthiobiotin in one elution steps. The purification was carried out at a constant flow of 0.75mL/min. The collected fractions are highlighted above (11-17).

A large amount of protein was eluted in a few fractions. Doing the purification step with the column upside down allowed the concentration of 35toDTG-Strep in only 7 fractions.

Samples from the elution step were analyzed by SDS-PAGE stained with Coomassie Blue. These samples were also analyzed by SDS-PAGE followed by Western blot. Anti-Strep-tag (1:2000) was used to identify 35toDTG-Strep_NP in the blot (Fig. 26).

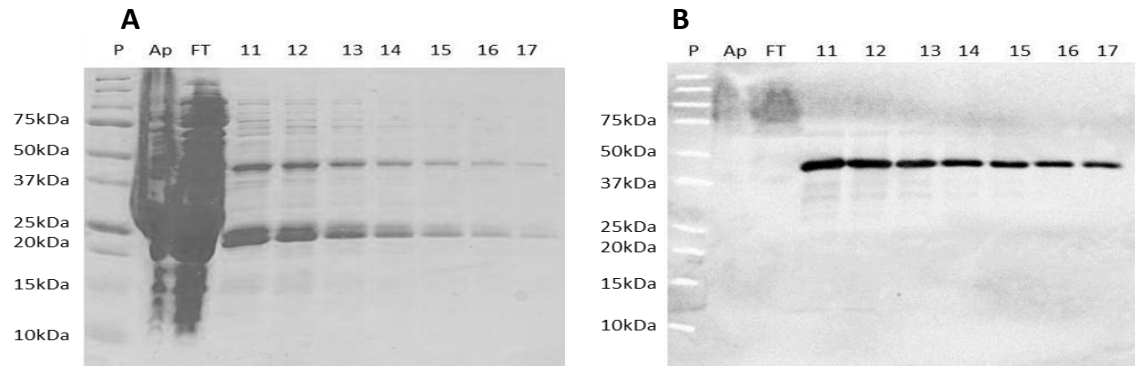

Figure 26. SDS-PAGE [A] and Western Blot [B] analysis of the factions collected from 35toDTG-Strep purification by Strep Trap affinity chromatography. The numbers correspond to the fraction number from the chromatogram shown in Figure 25; AP stands for the total soluble extract applied and FT to the protein not bounded to the column. SDS-PAGE was stained with Coomassie Blue. Antibody anti-Strep-tag (1:2000) was used to identify 35toDTG-Strap in the Western blot.

Although 35toDTG-Strep is also co-purified with other proteins, the amount of purified protein increased significantly, when compared to the protein purified by IMAC-Ni^{2+} (Fig10, 15, 21).

The proteolytic activity of fractions 11,12,13 collected from the StrepTrap affinity chromatography were screened against the typical aspartic protease substrate (MCA)-K-K-P-A-E-F-F-A-L-K-(DNP) in 50 mM sodium acetate buffer pH 4.0, 100 mM NaCl (Fig. 27). The fractions showed some activity with this substrate. Also, fractions with higher amount of protein present higher proteolytic activity, as well.

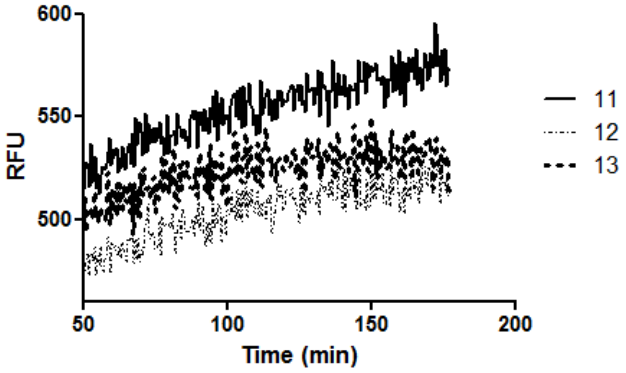

Figure 27. Proteolytic activity screening of recombinant 35toDTG-Strep-enriched fractions purified with StrepTrap affinity chromatography. The activity of 35toDTG-Strep was assayed using the typical AP substrate [(MCA)-K-K-P-A-E-F-F-A-L-K-(DNP)] in 50 mM sodium acetate buffer pH 4.0, 100 mM NaCl. The reaction was performed at 37°C for 175 min and negative control were made using only sample buffer, substrate and the assay buffers. The numbers correspond to the fractions analyzed in Fig. 26.

These results suggest that the Strep-tag may be a good strategy to further develop an alternative purification approach to obtain 35toDTG as a soluble protein.

3.3 Expression and Purification of 35toDTG-His as Inclusion Bodies

Given the low yields of protein obtained from the soluble fraction for all tested strategies, we decided to evaluate the effectiveness of producing the protein as inclusion bodies and subsequent refolding as an alternative strategy to obtain higher amounts of pure protein.

The expression screening results showed that expressing 35toDTG-His in BL21star in LB at 37°C for 3h produces a great amount of inclusion bodies with higher IPTG concentrations (Fig.8). So, to produce a large amount of inclusion bodies transformed BL21 star cells with 35toDTG-His pET23a construct were pre-inoculated in an Erlenmeyer flask with ZBAmp medium, at 30°C with constant rotation overnight. The culture was then diluted in 1L LBAmp Fernbach flasks (1:50 v/v). The cells were incubated at 37°C with constant rotation to an OD_{600nm} of 0.7 and then induced with 0.5 mM IPTG (Simões et al. 2007). The culture was incubated at 37°C with constant rotation for 3h.

For harvesting the cells, 1L cultures were centrifuged at 6000g for 20 min at 4°C.

The culture pellet was resuspended in 50 mMTris/HCl buffer pH 7.5 and 50mM NaCl (TN buffer) and then was frozen at -20°C.

3.3.1 Evaluation of 35toDTG-His refolding conditions

After thawing the cells, they suffer cell lysis and the inclusion bodies were washed in TN buffer overnight, followed by a second washing step overnight in TNT buffer. Then the inclusion bodies were centrifuged dissolved in 8 M Urea buffer with 7 mL/L β-mercaptoethanol.

The protein was then refolded by two different strategies.

In the first strategy the protein was refolded by rapid dilution (20-fold) into 20 mM Tris solution and the pH value was adjusted to pH8.0 (1st day: pH 9.0; 2^{sd} day pH 8.5; 3rd day: pH 8.0). The refolded protein was left at 4°C without agitation for one month.

Samples of the refolded protein were then evaluated by analytical size-exclusion chromatography. The protein was injected in a Superdex 200 HR 10/30 column (GE Healthcare), pre-equilibrated with 50 mM sodium acetate buffer pH 7.4, 100 mM NaCl, using a Prominence Shimadzu HPLC system (Fig28).

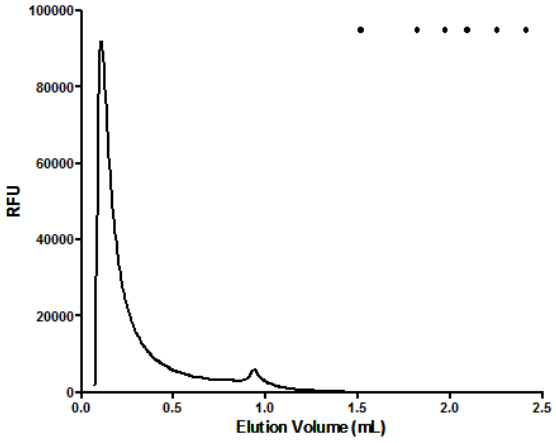

Figure 28.Analytical size-exclusion chromatography of recombinant 35toDTG-His refolded by rapid dilution for one month. The samples were injected in a Superdex 200 HR 10/30 pre-equilibrated with 50 mM sodium acetate buffer pH 7.4, 100 mM NaCl. The column was calibrated with protein standards (**dots**) in the same buffer: (from the left to the right: ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa) and ribonuclease (13.7 kDa).

As it can be seen the protein is completely aggregated, showing that this method was not useful to refold 35toDTG-His. So, a new approach was used to refold 35toDTG-His.

In refolding by rapid dilution, the denaturants concentration decreases drastically before the local high concentration of denatured protein is reduced. Under this conditions and with a high initial protein concentration, usually used in these methods, proteins easily aggregates (Yamaguchi et al. 2013). On the other hand, in a dialysis refolding method denaturant concentration decreases slowly, promoting formation of intra- or inter-molecular interactions and bonding again in a slower away, favoring correct folding (Yamaguchi et al. 2013). Unlike rapid dilution, refolding by dialysis doesn't require high protein concentration, being the initial and final concentration almost equal. Also it doesn´t lead to the accumulation of high concentration of denatured protein, avoiding aggregation (Yamaguchi et al. 2013).

So, refolding by dialysis seemed to be an attractive method to refold 35toDTG-His. We used a refolding concentration of 5 mg/mL of protein in 10 mL 8 M Urea buffer containing 10 mM DTT, 1 mM GSH, 0.1 mM GSSG and 7 mL/L β-mercaptoethanol. The sample was then diluted (10 fold) to a concentration of 0.5 mg/mL in 8M Urea buffer and then the protein was refolded by dialysis and the pH value was adjusted to pH 8.0 (1^{st} day: 20 mMTris/HCl buffer pH9.0; 2^{nd} day: 20 mM Tris/HCl buffer pH 8.5, 0.4M Urea; 3^{rd} day: 20 mM Tris/HCl buffer pH 8.0, 0.4M Urea). The refolded inclusion bodies were then left at 4°C without agitation for one month (Fig.29).

The chromatogram showed a high peak around 43kDa, the theoretical molecular weight of 35toDTG-His, corresponding to the monomeric form of the protein. A huge amount of protein with higher molecular weight is also present.

These other oligomeric forms may be just aggregated forms of 35toDTG-His. However, there wouldn't be the first time an atypical aspartic protease presents a different oligomeric state than the classic monomeric active state. CDR1, from *Arabidopsis thaliana*, require the formation of a homodimer for optimal catalytic activity (Simões et al. 2007). Although this hypothesis cannot be excluded, the aim of the evaluation of refolding conditions of 35toDTG-His was to find the best conditions to have a monomeric state.

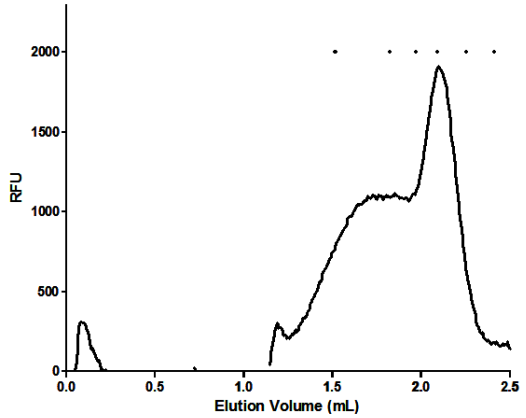

Figure 29.Analytical size-exclusion chromatography of recombinant 35toDTG-His refolded by dialysis for one month. The samples were injected in a Superdex 200 HR 10/30 pre-equilibrated with 50 mM sodium acetate buffer pH 7.4, 100 mM NaCl. The column was calibrated with protein standards (dots) in the same buffer (from the left to the right: ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa) and ribonuclease (13.7 kDa).

Therefore, refolding by dialysis proved to be a good method to obtain the protein in this oligomeric state, in high yields.

To evaluate the time needed to refold the protein, 35toDTG-His inclusion bodies were refolded by dialysis as described above and then left at 4°C without agitation only for 72h (Fig.30).

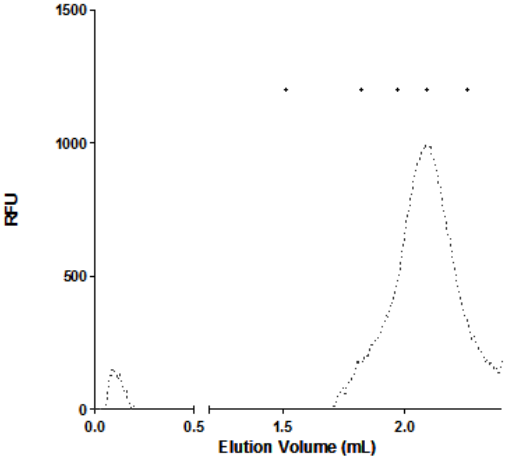

Figure 30. Analytical size-exclusion chromatography of recombinant 35toDTG-His refolded by dialysis for 72h. The samples were injected in a Superdex 200 HR 10/30 pre-equilibrated with 50 mM sodium acetate buffer pH 7.4, 100mM NaCl. The column was calibrated with protein standards (**dots**) in the same buffer (from the left to the right: ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa) and carbonic anhydrase (29 kDa).

The results proved that 72h are enough to make the protein assume a monomeric state. Looking closely to 1 month refolding and 72 refolding process (Fig.29 and Fig.30), the data suggests that 35toDTG-His began to aggregate if long refolding times are used.

The refolding by dialysis for 72 h showed to be the best condition to promote refolding of 35toDTG-His inclusion bodies, and therefore this condition was used to proceed to optimization of protein purification.

3.3.2 Recombinant 35toDTG-His purification with size-exclusion chromatography and an anionic exchange chromatography

The protein refolded by dialysis for 72h was concentrated using an Amicon Ultra Centrifugal Filter Unit concentrator (Stirred Cell 8200 –Millipore) and, after ultracentrigation, the sample was applied to a size exclusion chromatography Superdex200 prep grade (GE Healthcare), equilibrated with 20 mM Tris/HCl buffer pH 8.0, 0.4 M Urea. The purification step was carried out using a BioLogic DuoFlow FPLC system (Bio-Rad) at a constant flow of 1 mL/min. Fractions of 10 mL were collected (Fig.31).

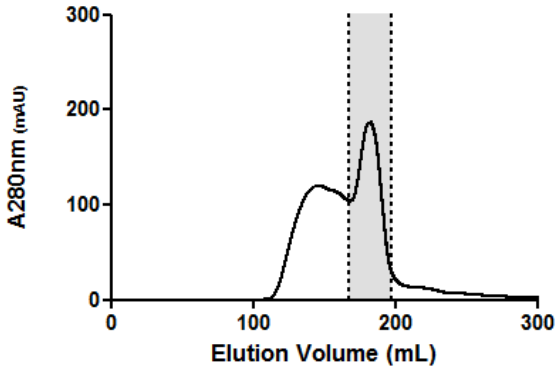

Figure 31. Size-exclusion chromatography of recombinant 35toDTG-His refolded by dialysis for 72h. The concentrated refolded protein was applied to a Superdex 200 prep grade pre-equilibrated with 20 mM Tris/HCl buffer pH 8.0, 0.4 M Urea. The purification step was carried out in a Bio-Rad FPLC system at a constant flow of 1 mL/min. Fractions in the grey area were applied to an anionic exchange chromatography.

One shoulder and one major peak are visible in the chromatogram. The first "shoulder" peak should correspond to a heterologous population of different oligomeric states. The other peak, with a maximum around 180 mL, should correspond to 35toDTG-His monomeric from.

Fractions corresponding to the monomeric 35toDTG-His were then applied to an anionic exchange Mono-Q 5/50 GL column (Amersham Biosciences). The column was equilibrated and washed with 20 mM Tris/HCl buffer pH 8.0, 0.4M Urea and the sample was eluted with a linear gradient of NaCl (0 to 1 M of NaCl) in the same buffer. The purification step was carried out using a BioLogic Duoflow FPLC system (Bio-Rad) at a constant flow of 0.75 mL/min. Fractions of 1 mL were collected through the elution steps (Fig.32).

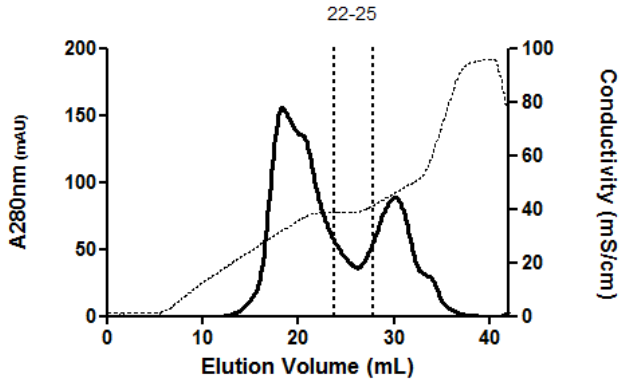

Figure 32. Purification of recombinant 35toDTG-His polled fractions from Superdex 200 by anionic exchange chromatography. The column was previously equilibrated with 20 mM Tris/HCl, 0.4 M, pH8.0 and washed with the same buffer. The protein was eluted with a linear gradient of NaCl (0 to 1 M of NaCl). The purification was carried out at a constant flow of 0.75mL/min. Collected fractions displaying enzymatic activity are highlighted (22-25).

Expression and purification of 35toDTG-His from inclusion bodies have a higher yield of purified protein than any other purification strategy used so far to obtain 35toDTG-His. Among atypical aspartic proteases described so far, excluding those which were purified from plant extracts, CDR1 and PCS1 were obtained by heterologous expression as inclusion bodies in BL21star, with similar conditions that those used in this work. To my knowledge, only OsCDR1 a rice atypical aspartic protease orthologous of *Arabidopsis thaliana*'s CDR1, have been expressed as soluble protein. However, this

protease was expressed on a small-scale system (5 mL of cells were harvested to obtain 1 mL of total protein). Even so the authors manage to a biochemical characterization of the protease (Prasad et al. 2010).

These results suggest that expression of atypical aspartic proteases as soluble protein in large-scale seems to promote its aggregation. It is tempting to suggest that a large-scale expression promote the increase of the insoluble fraction, leading to aggregation of the protein and, consequently, low yield of purified protein.

After the anionic exchange chromatography, several fractions were screened for proteolytic activity against the typical aspartic protease substrate (MCA)-K-K-P-A-E-F-F-A-L-K-(DNP) in 50 mM sodium acetate buffer pH 4.0 with 100 mM NaCl, and activity was only observed in fractions 22-25 (data not shown).

Fractions were then analyzed by SDS_PAGE, stained with Coomassie Blue, and by Western blot. Anti-His tag (1:10000) and an antibody specifically raised towards NP_181826 (1:20000) were used to identify 35toDTG-His (Fig.33).

As the results show, with this approach a high amount of protein was obtained, allowing a biochemical purification of 35toDTG-His. The fractions obtained are very clean; the protein is not co-purified with other contaminants.

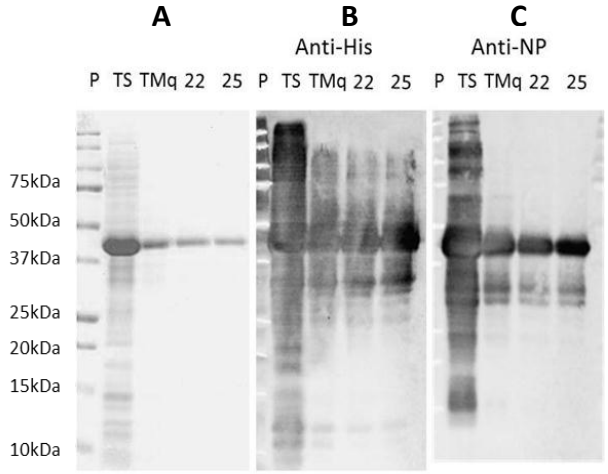

Figure 33. SDS-PAGE [A] and Western Blot [B, C] analysis of recombinant 35toDTG-His fractions collected from anionic exchange chromatography. The numbers correspond to the fraction number from the chromatogram shown in Fig.32; TS stands for the total soluble protein applied in Superdex 200 prep grad; TMq stands for the total soluble protein applied to the Mono-Q 5/50 GL column; Anti-His [B]: Western blot where 35toDTG-His was detected with an Anti-His tag antibody (1:10000); and Anti-NP [C]: Western blot where 35toDTG-His was detected with an Anti-NP_181826 antibody (1:20000).

The presence of 35toDTG-His was confirmed by the use of an Anti-His tag antibody and an Anti-NP_181826 antibody at approximately 43kDa in both blots.
The amount of purified protein was enough to do a biochemical characterization of 35toDTG-His.

3.4 Biochemical characterization of recombinant 35toDTG-His

As described before, the prosegment domain is involved in inactivation, correct folding, stability or intracellular sorting of proenzymes (Simões & Faro 2004).
Generally plant aspartic proteases are synthetized as inactive zymogenes. In order to become mature and active plant aspartic proteases, these preproproteases suffer a cleavage of the signal peptide, after translocation to the ER, followed by the cleavage of the prosegment (Simões & Faro 2004).
The amplicon 35toDTG, expressed in the work, was expressed without its signal peptide and prosegment domain. However, as showed before, 35toDTG had proteolytic activity against the typical aspartic protease substrate.
These suggest that the prosegment may have a different function in NP_181826 since it is not required to obtain a correct refolded structure and activity.
This is not the first time an aspartic protease prosegment showed an unusual function. Just recently Almeida and co-workers demonstrate that chlapsin, a very unique typical chloroplastidial aspartic protease from the green algae *Chlamydomonas reinhardtii*, also had proteolytic activity, when it was expressed without the signal peptide and the prosegment (Almeida et al. 2012). *Arabidopsis* CDR1 also showed protease activity without irreversible removal of its prosegment (Simões et al. 2007).

Previously, 35toDTG showed proteolytic activity against a fluorogenic typical aspartic protease substrate (MCA)-K-K-P-A-E-F-F-A-L-K-(DNP). This typical substrate presents a Phe-Phe bond, one of the most common cleavage sites that aspartic proteases show preferential specificity for (Vairo Cavalli et al. 2013). Fluorogenic substrates have in opposed terminal-ends a fluorophore and a quencher. When the substrate is cleaved, the fluorophore is no longer under a quenching effect, being able to emit fluorescence.

The fluorescence emitted per time unit is proportional to the proteolytic activity of the enzyme.

Having a substrate cleaved by 35toDTG-His, allows us to study the biochemical properties of the protease.

3.4.1 Effect of pH in proteolytic activity of recombinant 35toDTG-His

The effect of pH on 35toDTG-His was again tested towards the typical aspartic protease substrate (MCA)-K-K-P-A-E-F-F-A-L-K-(DNP). The assay was carried out in 50 mM sodium acetate buffer, 100mM NaCl at pH 3.0 to pH 6.0 or 50 mM sodium citrate buffer, 100 mM NaCl buffer at pH 2.5. The reaction was read by a Spectramax Gemini EM fluorimeter (Molecular Devices) at 37°C for 3h (Fig. 34).

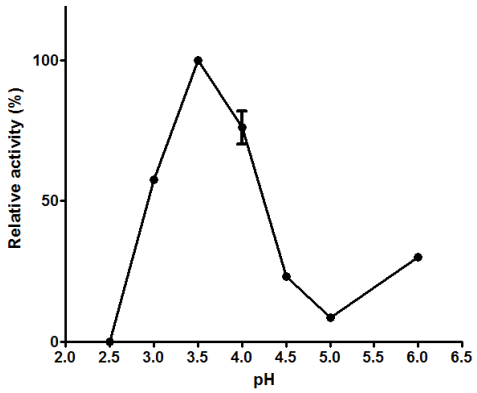

Figure 34. pH profile of recombinant 35toDTG-His purified by size exclusion chromatography and anionic exchange chromatography after refolding by dialysis. The effect of pH in 35toDTG-His activity was assayed using the typical AP substrate [(MCA)-K-K-P-A-E-F-F-A-L-K-(DNP)] and several buffers with different pH values (50mM sodium acetate buffer, 100 mM NaCl at pH 3.0-6.0 or 50 mM sodium citrate buffer, 100 mM NaCl at pH 2.5). The reaction was performed at 37°C for 3h and negative controls were made for each pH value using only sample buffer, substrate and the assay buffers.

Recombinant 35toDTG-His showed an optimum pH value at pH3.5 with the typical aspartic protease substrate. This pH curved was not significantly different from what was observed with 35toDTG-His expressed as soluble protein (Fig.13 and Fig.19). Although with the previous purification strategies the pH optimum for 35toDTG-His

with this substrate was pH 4.0, proteolytic activity at pH 3.5 was very close to the maximum activity at pH4.0.

A1 aspartic proteases, namely plant aspartic proteases present an acidic optimum pH value, just like 35toDTG-His. However, a pH 3.5 value seems to be a very low pH value that is not present in the chloroplast at physiological conditions. This suggests that 35toDTG-His may be present in stressful conditions to the plant, hypothesizing that 35toDTG-His may have a specialized function during stress, like CND41 (Murakami et al. 2000).

Although no major protein contaminants were visible in the SDS-PAGE gel of recombinant 35toDTG-His purification with an anionic exchange chromatography (Fig.32), proteolytic activity was again observed at pH 6.0.

Previous results obtained with the soluble protein suggested that this activity could be due to the presence of a co-purified protein from *E. coli*. Although this may also be the case in this purification strategy, this protease has to be stable enough to go through the refolding process and retain its proteolytic activity.

3.4.2 Inhibition profile of recombinant 35toDTG-His

In order to further characterize 35toDTG-His, the influence of several inhibitors in recombinant 35toDTG-His proteolytic activity was tested.

Activity was tested against the typical aspartic protease substrate in the presence of Pepstatin, a typical aspartic protease inhibitor, EDTA, a metalloprotease inhibitor due to its chelating characteristic, E-64, an inhibitor of cysteine proteases, Pefabloc, an irreversible serine protease inhibitor, and Bestatin, a metalloprotease inhibitor, specifically for aminopeptidases.

The protease was incubated with the different inhibitors for 5 minutes at room temperature. The reaction was carried out at a Spectramax Gemini EM fluorimeter (Molecular Devices) set to λ_{ex}=328nm and to λ_{em}=393nm, 37°C for 3 h.

The higher amount of cysteines present in 35toDTG-His suggests that the redox environment may be crucial for the protease stability and activity. Therefore, several redox agents were also tested against 35toDTG-His.

Also, due to the putative chloroplastidial localization of NP_181826, the effect of several nucleotides may be extremely important to NP_181826 proteolytic activity. So, their effect on 35toDTG-His activity was also tested in the optimum proteolytic activity conditions of 35toDTG-His (Fig.35).

The proteolytic activity of 35toDTG-His was completely inhibited be Pepstatin A, validating 35toDTG-His as an aspartic protease. However, Bestatin also had a strong influence in the protease activity, as well as EDTA, suggesting that the protein may have some kind of metalloprotease activity, namely aminopeptidases activity.

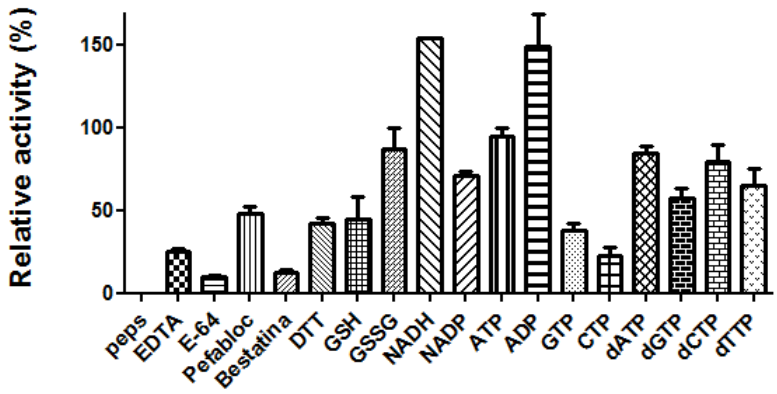

Figure 35. Inhibition profile of recombinant 35toDTG-His purified size exclusion chromatography and anionic exchange chromatography after refolding by dialysis. The effect of several inhibitors (final concentration and solvents in Table 3) in 35toDTG-His activity was assayed using the typical AP substrate [(MCA)-K-K-P-A-E-F-F-A-L-K-(DNP)] and 50 mM sodium acetate buffer, 100mM NaCl at pH 3.5. The inhibitors were pre-incubated for 5 minutes with the protease and the reaction was performed at 37°C for 3h. The activity of the control sample, without any inhibitor, was taken as 100%. A negative control was made using only sample buffer, substrate and the assay buffer. For each inhibitor, the correspondent solvent was used as a negative control; the inhibition of each inhibitor was calculated considering the inhibition effect of the solvent.

Unexpectedly, E64, an irreversible cysteine protease inhibitor, seems to have a huge inhibitory effect on this protease. To my knowledge, this is the first time an inhibition effect of E64 on atypical aspartic proteases is described.

When tested against redox agents, DTT and GSH had some influence on 35toDTG-His activity compared to GSSG's no effect, supporting the idea that the redox environment may be crucial to the stability and activity of the enzyme. 35toDTG-His also suffer some inhibition in the presence of GTP and CTP.

NADH and ADP effect on 35toDTG-His was again extremely interesting. These two agents have an activation effect on the protease, increasing its activity around 50%. The results suggest that the rate ATP/ADP present in the chloroplast may have a key influence in the protease activity. The increase of ADP, can lead to 35toDTG-His activation.

Metabolism in the chloroplast is largely influenced by these nucleotides and redox agents (Almeida et al. 2012). The inhibition effect of DTT, GSH, GTP and CTP and the strong activation of 35toDTG-His in the presence of NADH and ADP suggests that the chloroplast environment and metabolites influence directly the protease activity.

3.4.3 Proteolytic Specificity of recombinant 35toDTG-His

In general, aspartic proteases show preferential specificity for cleavage between hydrophobic residues (e.g. Phe, Val, Ile, Leu) on either side of the scissile bond, at P1 and P1' positions. Phe-Phe is one of the most common cleavage sites (Vairo Cavalli et al. 2013). Atypical aspartic proteases show a very different range of cleavage specificity, related to their biological role in the organisms and to their natural substrates.

CDR1 has extremely restricted cleavage specificity, being able to cleave Leu-Tyr bond of oxidized insulin β-chain, at pH6.0. This protease is involved in the proteolytic generation of a specific peptide elicitor that may activation of salicylic-acid-dependent local and systemic defense responses (Xia et al. 2004). On the other hand Nepenthesin I show a broader specificity, being able to cleave the oxidized insulin β-chain in more than five different places, at pH3.0 (Athauda et al. 2004). This broad specificity is in agreement with its function as a recycling protein, used to obtain nitrogen from the plants preys. Also results suggest that CND41 may have a broad specificity, being able to act on different denatured chloroplast proteins, as well(Kato et al. 2004). Several reports show that the proteolytic activity by CND41 is the source of amino acids being transported to developing organs, in senescence conditions (Kato et al. 2004).

To start evaluating 35toDTG-His cleavage specificity, the protease was incubated with oxidized insulin β-chain in 100 mM sodium acetate buffer pH 3.5. Digestion was carried out for 24h at 37°C. After stopping the reaction by adding 0.6% TFA the resulting soluble fragments were separated by reverse-phase HPLC in a KROMASIL C18 column (Teknokroma) using a Prominence Shimadzu HPLC system, pre-equilibrated with 0.1% TFA. The elution was carried out by a linear gradient of acetonitrile (0 to 80%) in 0.1% TFA at a flow rate of 1 mL/min. Peptide elution was monitored at 220nm (Fig.36).

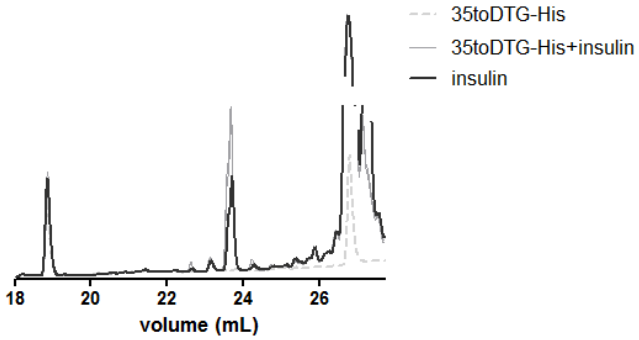

Figure 36. Reversed-phase HPLC elution profiles of digested oxidized insulin β-chain by recombinant 35toDTG-His. Oxidized insulin β-chain was incubated with recombinant 35toDTG-His for 24 h at 37°C, in 100 mM sodium acetate buffer pH3.5. After precipitation with 0.6% TFA, the soluble fragments were injected in a reverse-phase HPLC column equilibrated with 0.1% TFA. The peptides were eluted with an increasing linear gradient of acetonitrile (0 to 80%) in the same buffer. Negative controls of the Oxidized insulin β-chain and 35toDTG-His fraction were performed by incubating them, separately, in 100 mM sodium acetate buffer pH3.5 for 24h at 37°C. 35toDTG-His stands for the negative control with the sample, 35toDTG-His+insulin stands for the digested insulin fragments; insulin stands for the negative control with insulin. The peak resulting from digestion is highlighted with one arrow.

The digestion profile of oxidized insulin β-chain only presents one peak (around 24 mL). The results suggest that 35toDTG-His has extremely restricted cleavage specificity. Considering other atypical aspartic proteases, it is tempting to suggest that 35toDTG-His may have a very specific natural substrate and that it may be involved in a very specific function. The identification of the cleavage site is being processed using Mass Spectrometry in the Proteomics unit at CNC/Biocant.

3.5 Recombinant 35toATG

Although plant aspartic proteases are very different among them, some general features are commonly used to describe them.

The catalytic activity of plant aspartic proteases is based on the interaction of two aspartic residues. These two aspartic residues are responsible for electrons transfer during a nucleophilic attack of a water molecule, polarized by the aspartic residue of the second triad.

Also, in general, proteolytic activity of aspartic proteases, namely plant aspartic proteases, experience inhibition in the presence of Pepstatin A (Faro & Gal 2005; Simões & Faro 2004). In fact, this inhibitor had been widely used to identify and purify aspartic proteases (Athauda et al. 2004).

Atypical aspartic proteases, as described above, are very different from other plant aspartic proteases and even among them. They have very different sensitivities to Pepstatin's effect. Although ASPG1 and Nepenthesin I and II suffer a strong inhibition with Pepstatin A, PCS1 and CND41 proteolytic activities are not affected by this inhibitor (Athauda et al. 2004; Yao et al. 2012; Silva 2008; Murakami et al. 2000). Thus, the proteolytic activity inhibition by Pepstatin A is not enough to classify one protease as an aspartic protease.

To validate 35toDTG-His as an aspartic protease, an active-site mutant was created. Its aspartic residue of the first catalytic triad was replaced by an alanine residue by site-directed mutagenesis (D178A full sequence numbering). The 35toATG active-site mutant was generated using the QuikChange^{TM} Site-Directed Mutagenesis Kit (Stratagene). The expression vector pET23a with 35toDTG-His insert was used as DNA template and it was amplified using specific primers.

3.5.1 Expression and Purification of recombinant 35toATG

Transformed BL21 star cells with 35toDTG-His active-site mutant pET23a construct were inoculated in 1L LBAmp. Expression and isolation of the inclusion bodies was

carried out as described for wild-type 35toDTG-His. The protein was refolded by dialysis for 72h, as described for wild-type 35toDTG-His.

The soluble refolded protein, after concentration, was applied to a size exclusion chromatography Superdex200 prep grade (GE Healthcare), equilibrated with 20mM Tris/HCl, 0.4M Urea, pH 8.0. The purification step was carried out using an AKTA FPLC system (Ge Healthcare) at a constant flow of 1 mL/min. Fractions of 10 mL were collected (Fig.37).

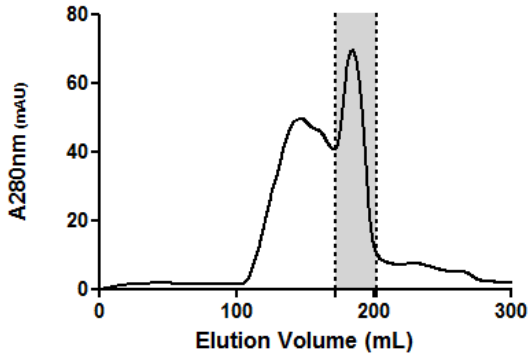

Figure 37. Size-exclusion chromatography of recombinant 35toATGrefolded by dialysis for 72h. The concentrated refolded protein was applied to a Superdex 200 prep grade pre-equilibrated with 20mM Tris/HCl buffer pH 8.0, 0.4 M Urea. The purification step was carried out in an AKTA FPLC system at a constant flow of 1 mL/min. Fractions in the grey area were applied to an anionic exchange chromatography.

Fractions corresponding to the monomeric 35toATG active-site mutant were applied to an anionic exchange Mono-Q 5/50 GL column (Amersham Biosciences). The column was equilibrated and washed with 20mM Tris/HCl buffer pH 8.0, 0,4 M Urea and the sample was eluted with a linear gradient of NaCl (0 to 1 M of NaCl) in the same buffer. The purification step was carried out using an AKTA FPLC system (Ge Healthcare) at a constant flow of 0.75 mL/min. Fractions of 1 mL were collected through the elution steps (Fig.38).

Comparing chromatogram profiles from 35toDTG-His and from its active-site mutant anionic exchange chromatographies (Fig.32 and Fig.38), they seem very similar to each other.

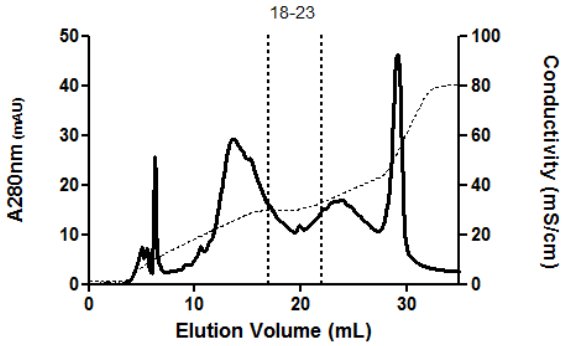

Figure 38. Purification of recombinant 35toATG polled fractions from Superdex 200 by anionic exchange chromatography. The column was previously equilibrated with 20 mM Tris/HCl, 0.4 M Urea, pH8.0 and washed with the same buffer. The protein was eluted with a linear gradient of NaCl (0 to 1 M of NaCl). The purification was carried out at a constant flow of 0.75 mL/min. Collected fractions are highlighted in grey.

It was possible to delimit the fractions corresponding to those where proteolytic activity was detected in wild-type 35toDTG-His purification (see Fig.32).

Fractions were then analyzed by SDS_PAGE, stained with Coomassie Blue, and by Western blot. Anti-His tag (1:10000) was used to identify 35toATG (Fig.39).

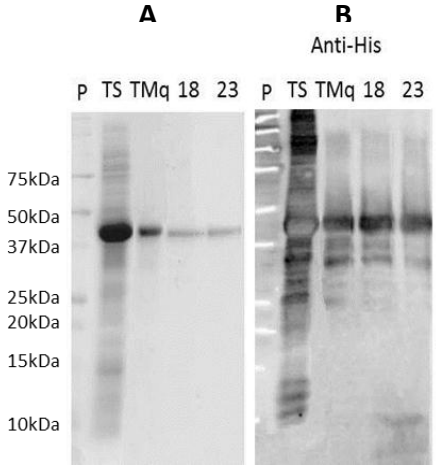

Figure 39. SDS-PAGE [A] and Western Blot [B] analysis of the factions collected from anionic exchange chromatography. The numbers correspond to the fraction number from the chromatogram shown in Fig.38; TS stands for the total soluble protein applied to the Superdex 200 prep grad; TMq stands to the total soluble protein applied to the Mono-Q 5/50 GL column; Anti-His tag antibody (1:10000) was used to detect 35toATG.

As the results show, a high amount of protein was obtained. The fractions are very clean; the protein is not co-purified with other contaminants.

The presence of 35toATG was confirmed by the use of an Anti-His tag antibody at approximately 43kDa.

The activity of the collected fractions during the 35toATG purification were analyzed against the typical aspartic protease subtract (MCA)-K-K-P-A-E-F-F-A-L-K-(DNP). The assay was carried out in 50 mM sodium acetate buffer pH 3.5, 100 mM NaCl and the reaction was read by a Spectramax Gemini EM fluorimeter (Molecular Devices) at 37°C for 3h. The same amount of recombinant wild-type and mutant 35toATG protein was added in all assays and no proteolytic activity was observed for the active-site mutant. The results confirm recombinant 35toDTG as an active aspartic protease.

4. Conclusion remarks and Future Prospects

In this work we tried to optimize the expression and purification of the atypical aspartic protease NP_181826 and perform its biochemical characterization. The work was mainly focused on the construct 35toDTG, which corresponds to a putative mature form of the enzyme, starting 35 amino acids before the first catalytic triad (DTG).

Several strategies were used to purify 35toDTG fused to a His-Tag as a soluble protein. However, the protein was always co-purified with other proteins. Although oxidases were pre-expressed together with the protein of interest, this *stickiness* character may suggest that the protein was not folded correctly and that to maintain a soluble form it interacts with proteins present in *E. coli*. Therefore, expressing the protein in the soluble form proved to be not the best strategy to obtain 35toDTG. It resulted in very low amounts of protein and even the use of Ultra Yield flasks resulted in very poor yields of purified protein.

Purification of 35toDTG in frame with a Strep-tag proved to be a better approach to obtain higher yields of purified protein, though it was still co-purified with other contaminants. This could be an interesting first step to develop a new purification strategy of 35toDTG as a soluble protein.

Among atypical aspartic proteases described so far two of them were obtained by heterologous expression as inclusion bodies. Refolding protein from inclusion bodies isn't always a straightforward process. Not all the proteins refold in the same way, which require an optimization of the refolding conditions. Refolding by dialysis proved to be the best way of refolding 35toDTG. Also, with only 72h of refolding time it was possible to obtain less aggregated protein. With longer times of refolding, the protease tended to aggregate. Thus, expression of 35toDTG as inclusion bodies and refolding by dialysis resulted in higher yields of purified protein that allowed its biochemical characterization.

Recombinant 35toDTG-His showed proteolytic activity against a fluorogenic typical aspartic protease substrate, (MCA)-K-K-P-A-E-F-F-A-L-K-(DNP), with an optimum pH value of 3.5. It is not likely that this low pH value occurs in the chloroplast under physiological conditions, suggesting that 35toDTG-His may be activated in stress conditions.

Although, the recombinant 35toDTG-His was expressed without its signal peptide and prosegment domain, 35toDTG had proteolytic activity against the typical aspartic protease substrate. These results demonstrate that the prosegment was not required to obtain a correct structure and activity, suggesting that NP181826 prosegment domain may have a different function.

The digestion profile of oxidized insulin β-chain only presents one cleavage site. The results suggest that 35toDTG has extremely restricted cleavage specificity.
It is tempting to suggest that 35toDTG may have a very specific natural substrate and that it may be involved in a very specific function.
The proteolytic activity of 35toDTG was completely inhibited by Pepstatin A, validating 35toDTG as an aspartic protease. This atypical aspartic protease was also inhibited by Bestatin as well as EDTA, suggesting that it can have some kind of metalloprotease activity, namely aminopeptidase activity.
The redox environment and the equilibrium of nucleotide species are extremely important to chloroplast metabolism. Recombinant 35toDTG showed some inhibition in the presence of DTT and GSH, with no effect in the presence of GSSG. This supports the idea that redox environment may be crucial to the stability and proteolytic activity of the protease. Also, NADH and ADP effect on 35toDTG was again extremely interesting. These two agents have an activation effect on the protease, increasing its activity around 50%. The results suggest that the rate ATP/ADP present in the chloroplast may have a key influence in the protease activity.
Complete inhibition in the presence of Pepstatin A and total loss of the proteolytic activity by 35toATG validate NP_181826 as an aspartic protease.

Although this work showed some interesting features of NP_181826, many questions remain unanswered.
Other expression systems, such as yeast or even insect cell or heterologous expression in plants, could be used to express the protease in the soluble form. These eukaryotic expression systems promote naturally the formation of disulfide bonds, leading likely to the expression of 35toDTG with the correct folding.

Additionally, it will be interesting to understand the effect of DNA in ΔSPTP construct in physiological and stress conditions. This construct may also give some experimental evidences of auto-processing of the protease.

Another important issue will be to determine NP_181826 natural interaction partners. Our experimental evidences suggest that the protease may have a very specific substrate and so a very specific function. Determination of its natural partners will provide information about its biological role in *Arabidopsis thaliana* chloroplast.

Regarding its bipartite presequence, it will be interesting to understand in what way the protease reaches the chloroplast and what is the relevance of the signal peptide in this targeting.

Bibliography

Almeida, C.M. et al., 2012. Chlapsin, a chloroplastidial aspartic proteinase from the green algae Chlamydomonas reinhardtii. *Planta*, 236, pp.283–296.

Athauda, S.B.P. et al., 2004. Enzymic and structural characterization of nepenthesin, a unique member of a novel subfamily of aspartic proteinases. *The Biochemical journal*, 381, pp.295–306.

Berkmen, M., 2012. Production of disulfide-bonded proteins in Escherichia coli. *Protein expression and purification*, 82, pp.240–251.

Bi, X., Khush, G.S. & Bennett, J., 2005. The rice nucellin gene ortholog OsAsp1 encodes an active aspartic protease without a plant-specific insert and is strongly expressed in early embryo. *Plant & cell physiology*, 46, pp.87–98.

Brunet, J. et al., 2009. Accumulation of lead in the roots of grass pea (Lathyrus sativus L.) plants triggers systemic variation in gene expression in the shoots. *Chemosphere*, 77, pp.1113–1120.

Bryksa, B.C. et al., 2011. Structure and mechanism of the saposin-like domain of a plant aspartic protease. *The Journal of biological chemistry*, 286, pp.28265–28275.

Chen, J. et al., 2008. A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility of indica-japonica hybrids in rice. *PNAS*, 105, pp.11436–11441.

Chen, J. et al., 2009. Aspartic proteases gene family in rice: Gene structure and expression, predicted protein features and phylogenetic relation. *Gene*, 442, pp.108–118.

Coates, L. et al., 2008. The catalytic mechanism of an aspartic proteinase explored with neutron and X-ray diffraction. *Journal of the American Chemical Society*, 130, pp.7235–7237.

Dunn, B.M., 2002. Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chemical reviews*, 102, pp.4431–4458.

Egas, C. et al., 2000. The saposin-like domain of the plant aspartic proteinase precursor is a potent inducer of vesicle leakage. *The Journal of biological chemistry*, 275, pp.38190–38196.

Faro, C. & Gal, S., 2005. Aspartic proteinase content of the Arabidopsis genome. *Current Protein and Peptide Science*, 6, pp.493–500.

Frazao, C., 1999. Crystal Structure of Cardosin A, a Glycosylated and Arg-Gly-Asp-containing Aspartic Proteinase from the Flowers of Cynara cardunculus L. *Journal of Biological Chemistry*, 274(39), pp.27694–27701.

Gao, Y. et al., 2011. Expression of aspartyl protease and C3HC4-type RING zinc finger genes are responsive to ascorbic acid in Arabidopsis thaliana. *Journal of experimental botany*, 62, pp.3647–3657.

Ge, X. et al., 2005. An Arabidopsis aspartic protease functions as an anti-cell-death component in reproduction and embryogenesis. *EMBO reports*, 6, pp.282–288.

Goodstadt, L. & Ponting, C.P., 2001. CHROMA: consensus-based colouring of multiple alignments for publication. *Bioinformatics (Oxford, England)*, 17, pp.845–846.

Grennan, A.K., 2006. Regulation of starch metabolism in Arabidopsis leaves. *Plant physiology*, 142, pp.1343–1345.

Gruber, A. et al., 2007. Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. *Plant molecular biology*, 64, pp.519–530.

Hatahet, F. et al., 2010. Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of E. coli. *Microbial cell factories*, 9, pp.67–78.

Ji, Q. et al., 2012. Molecular Basis Underlying the S5-Dependent Reproductive Isolation and Compatibility of Indica/Japonica Rice Hybrids. *Plant physiology*, 158, pp.1319–1328.

Kato, Y. et al., 2005. Post-translational regulation of CND41 protease activity in senescent tobacco leaves. *Planta*, 222, pp.643–651.

Kato, Y. et al., 2004. The DNA-binding protease, CND41, and the degradation of ribulose-1,5-bisphosphate carboxylase/oxygenase in senescent leaves of tobacco. *Planta*, 220, pp.97–104.

Kervinen, J. et al., 1999. Crystal structure of plant aspartic proteinase prophytepsin: inactivation and vacuolar targeting. *The EMBO journal*, 18(14), pp.3947–55.

Larkin, M. a et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, 23, pp.2947–2948.

Loake, G. & Grant, M., 2007. Salicylic acid in plant defence--the players and protagonists. *Current opinion in plant biology*, 10, pp.466–472.

Milisavljevic, M. et al., 2011. Cell wall localization of the aspartic proteinase from buckwheat (FeAPL1) over-expressed in tobacco BY-2 cells. *Journal of the Serbian Chemical Society*, 76, pp.1229–1236.

Milisavljevic, M. et al., 2007. Isolation and structural analysis of a gene coding for a novel type of aspartic proteinase from buckwheat seed (Fagopyrum esculentum Moench). *Archives of Biological Sciences*, 59, pp.119–124.

Milisavljevic, M.D. et al., 2008. Two types of aspartic proteinases from buckwheat seed--gene structure and expression analysis. *Journal of plant physiology*, 165, pp.983–990.

Muñoz, F.F. et al., 2010. The swaposin-like domain of potato aspartic protease (StAsp-PSI) exerts antimicrobial activity on plant and human pathogens. *Peptides*, 31, pp.777–785.

Murakami, S. et al., 2000. Protease activity of CND41, a chloroplast nucleoid DNA-binding protein, isolated from cultured tobacco cells. *FEBS letters*, 468, pp.15–18.

Nakano, T. et al., 1997. A novel protein with DNA binding activity from tobacco chloroplast nucleoids. *The Plant cell*, 9, pp.1673–1682.

Nguyen, V.D. et al., 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, pp.1–13.

Noodén, L.D., 2013. Defining Senescence and Death in Photosynthetic Tissues. In B. Biswal, K. Krupinska, & U. C. Biswal, eds. *Plastid Development in Leaves during Growth and Senescence, Advances in Photosynthesis and Respiration Vol36*. Dordrecht: Springer Netherlands, pp. 283–306.

Nozach, H. et al., 2013. High throughput screening identifies disulfide isomerase DsbC as a very efficient partner for recombinant expression of small disulfide-rich proteins in E. coli. *Microbial cell factories*, 12, pp.37–52.

Olivares, J.E. et al., 2011. Nodulin 41, a novel late nodulin of common bean with peptidase activity. *BMC plant biology*, 11, pp.134–147.

Paparelli, E. et al., 2012. Misexpression of a chloroplast aspartyl protease leads to severe growth defects and alters carbohydrate metabolism in Arabidopsis. *Plant physiology*, 160, pp.1237–1250.

Phan, H.A. et al., 2011. The MYB80 Transcription Factor Is Required for Pollen Development and the Regulation of Tapetal Programmed Cell Death in Arabidopsis thaliana. *The Plant cell*, 23, pp.2209–2224.

Ponting, C. & Russell, R., 1995. Swaposins: circular permutations within genes encoding saposin homologues. *Trends in biochemical sciences*, 20, pp.179–180.

Prasad, B.D. et al., 2010. Heterologous expression and characterization of recombinant OsCDR1, a rice aspartic proteinase involved in disease resistance. *Protein expression and purification*, 72, pp.169–174.

Prasad, B.D. et al., 2009. Overexpression of rice (Oryza sativa L.) OsCDR1 leads to constitutive activation of defense responses in rice and Arabidopsis. *Molecular plant-microbe interactions : MPMI*, 22, pp.1635–1644.

Radhamony, R.N. & Theg, S.M., 2006. Evidence for an ER to Golgi to chloroplast protein transport pathway. *Trends in cell biology*, 16, pp.385–387.

Rawlings, N.D., Barrett, A.J. & Bateman, A., 2012. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic acids research*, 40, pp.D343–D350.

Saibo, N.J.M., Lourenço, T. & Oliveira, M.M., 2009. Transcription factors and regulation of photosynthetic and related metabolism under environmental stresses. *Annals of botany*, 103, pp.609–623.

Schaller, A., 2004. A cut above the rest: the regulatory function of plant proteases. *Planta*, 220, pp.183–197.

Schmidt, T.G.M. & Skerra, A., 2007. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nature protocols*, 2, pp.1528–1535.

Silva, P., 2008. *Expressão, Purificação e Caracterização da Protease Aspártica Atípica Recombinante PCS1 de Arabidopsis thaliana*. Faculty of Sciences and Technology of University of Coimbra.

Simões, I. et al., 2007. Characterization of recombinant CDR1, an Arabidopsis aspartic proteinase involved in disease resistance. *The Journal of biological chemistry*, 282, pp.31358–31365.

Simões, I. & Faro, C., 2004. Structure and function of plant aspartic proteinases. *European journal of biochemistry*, 271, pp.2067–2075.

Törmäkangas, K. et al., 2001. A vacuolar sorting domain may also influence the way in which proteins leave the endoplasmic reticulum. *The Plant cell*, 13, pp.2021–2032.

Vairo Cavalli, S. et al., 2013. Properties and applications of phytepsins from thistle flowers. *Phytochemistry*, 92, pp.16–32.

Xia, Y. et al., 2004. An extracellular aspartic protease functions in Arabidopsis disease resistance signaling. *The EMBO journal*, 23, pp.980–988.

Xu, X. et al., 2010. Flower-specific expression of Arabidopsis PCS1 driven by AGAMOUS second intron in tobacco decreases the fertility of transgenic plants. *Molecular Breeding*, 27, pp.337–346.

Yamaguchi, S. et al., 2013. Protein refolding using chemical refolding additives. *Biotechnology journal*, 8, pp.17–31.

Yan, S.-P. et al., 2006. Comparative proteomic analysis provides new insights into chilling stress responses in rice. *Molecular & cellular proteomics*, 5, pp.484–496

Yao, X. et al., 2012. Overexpression of the aspartic protease ASPG1 gene confers drought avoidance in Arabidopsis. *Journal of experimental botany*, 63, pp.2579–2593.